

# Robust Inversion Methods for Aerosol Spectroscopy

Graham Alldredge<sup>a</sup> and Tobias Kyrion<sup>b</sup>

<sup>a</sup>Center for Computational Engineering Science, RWTH Aachen University, Aachen, Germany

<sup>b</sup>RWTH Aachen University, Aachen, Germany

January 17, 2017

## Abstract

The Fast Aerosol Spectrometer (FASP) is a device for spectral aerosol measurements. Its purpose is to safely monitor the atmosphere inside a reactor containment. First we describe the FASP and explain its basic physical laws. Then we introduce our reconstruction methods for aerosol particle size distributions designed for the FASP. We extend known existence results for constrained Tikhonov regularization by uniqueness criteria and use those to generate reasonable models for the size distributions. We apply a Bayesian model-selection framework on these pre-generated models. We compare our algorithm with classical inversion methods using simulated measurements. We then extend our reconstruction algorithm for two-component aerosols, so that we can simultaneously retrieve their particle-size distributions and unknown volume fractions of their two components. Finally we present the results of a numerical study for the extended algorithm.

## 1 The FASP measurement device

The FASP is an optical measurement device for aerosol particle size distributions in rigid environments where the temperature may surpass 200°C and the pressure 8 bar over atmospheric pressure, cf. [1, 2]. The aerosol particles themselves may be acidic as well. The FASP is split into a detector head and into a unit containing an evaluation computer and a light source with different adjustable light wavelengths. The sensitive evaluation and light source unit is connected with the robust detector head via two optical fibers.

The detector head is the only part of the FASP which extends into the containment with the aerosol to be measured and it consists of a pneumatically propelled tube. By moving the tube one can adjust a short or a long measurement path, where the two path lengths are 400 and 800 mm respectively. The sought-after aerosol particle size distributions are reconstructed from the light intensity loss on the gap distance between long and short path, so the FASP works in a similar way to a White cell. The detector head is equipped with two light detectors. The first one can receive light with wavelengths in the infrared domain from 0.8 - 3.4  $\mu\text{m}$ , and the other one in the visible domain from 0.5 - 0.8  $\mu\text{m}$ .



Figure 1.1: The detector head with the movable tube (source: [3])

The ends of the optical fibers have to be floated with protective gas to shield them from harmful aerosol particles. These particle-free sections have to be subtracted from the actual geometric path lengths. This is not problematic since this does not change the gap distance.

Let  $l$  denote a current light wavelength used in a measurement,  $G_{long}$  the geometric or unfloated long path and  $G_{short}$  the geometric short path. The section floated with protective gas is labeled with  $x$ . Then the true path lengths are given by  $L_{long} := G_{long} - x$  and  $L_{short} := G_{short} - x$ .

Let  $M_{long}(l)$  and  $M_{short}(l)$  be the measured intensities for long and short path, both perturbed by detector offsets  $O_{long}(l)$  and  $O_{short}(l)$  caused by ambient radiation.

Then the intensities cleaned from the detector offsets are given by  $I_{long}(l) := M_{long}(l) - O_{long}(l)$  and  $I_{short}(l) := M_{short}(l) - O_{short}(l)$ .

According to the law of Beer-Lambert we have the relation

$$I_{long}(l) = I_{short}(l) \exp \left( - (L_{long} - L_{short}) \int_0^\infty k(r, l) n(r) dr \right), \quad (1.1)$$

where  $n(r)$  is the sought-after unknown particle size distribution. The kernel function  $k(r, l) := \pi r^2 Q_{ext}(m_{med}(l), m_{part}(l), r, l)$  depends on both complex refractive indices  $m_{med}(l)$  and  $m_{part}(l)$  of the surrounding medium and the scattering aerosol particles which depend on the wavelength  $l$  of the incident light. The Mie extinction efficiency  $Q_{ext}(m_{med}(l), m_{part}(l), r, l)$  of a spherical particle with radius  $r$  illuminated by light with wavelength  $l$  is derived from the general solution to the corresponding boundary value problem for Maxwell's equations and was first introduced in the pioneering article [4]. We adopt the numerical approximation of the Mie extinction efficiency in an absorbing medium from [5]. From all of this follows

$$\int_0^\infty k(r, l) n(r) dr = e(l) \quad \text{with} \quad e(l) = - \frac{\log(I_{long}(l)) - \log(I_{short}(l))}{L_{long} - L_{short}}. \quad (1.2)$$

## 2 Modeling of FASP Measurement Data Inversions

Let the measurement data  $e(l)$  be an error-contaminated right-hand side for (1.2) and  $(Kn)(l) := \int_0^\infty k(r, l) n(r) dr$  the compact linear operator with unbounded inverse which maps possible size distributions  $n(r)$  to the left-hand side of (1.2). We wish to reconstruct  $n(r)$  from  $e(l)$  by inverting the equation

$$Kn = e. \quad (2.1)$$

Here and in the following we omit the dependence on  $r$  and  $l$  for better readability. We assume that  $e$  is given as a vector of finitely many independent Gaussian random variables  $e_i$  with standard deviations  $\sigma_i$  and means  $\mu_i$ , i.e.  $e_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ . In the framework of Bayesian inference these are our *observed random variables*. Now let  $\mathbf{n} \in \mathbb{R}^N$  be a discrete approximation to  $n$  and  $\mathbf{K}_N \in \mathbb{R}^{N_l \times N}$  the *kernel matrix* which correspondingly approximates the integral operator  $K$ . The details of these discretizations will be given in Section 5.1. We set up the covariance matrix  $\mathbf{\Sigma}_\sigma = \text{diag}(\sigma_1^2, \dots, \sigma_{N_l}^2)$ .

Then the *observed model uncertainty* under the assumption  $(\mathbf{K}_N \mathbf{n})_i = \mu_i$  obeys the probability distribution

$$p_{\text{observed}}(\mathbf{e}|\mathbf{n}) \propto \exp(-\frac{1}{2}\|\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}^{-\frac{1}{2}}(\mathbf{K}_N \mathbf{n} - \mathbf{e})\|_2^2). \quad (2.2)$$

After selecting a subjective *prior distribution*  $p_{\text{prior}}(\mathbf{n})$  which incorporates known *a priori* information about  $\mathbf{n}$  independent from the observed variable  $\mathbf{e}$  we use Bayes' rule to obtain the *posterior distribution*  $p_{\text{posterior}}(\mathbf{n}|\mathbf{e})$  with

$$p_{\text{posterior}}(\mathbf{n}|\mathbf{e}) \propto p_{\text{observed}}(\mathbf{e}|\mathbf{n}) \times p_{\text{prior}}(\mathbf{n}). \quad (2.3)$$

A more elaborate presentation of this Bayesian framework will be given in Section 4. By applying a Tikhonov prior distribution

$$p_{\text{prior}}(\mathbf{n}) \propto \exp(-\frac{1}{2}\gamma\|\mathbf{n}\|_2^2)I_S(\mathbf{n}),$$

where  $\gamma \geq 0$  is a regularization parameter and  $I_S(\mathbf{n})$  is the indicator function of the convex set

$$S := \{\mathbf{n} \in \mathbb{R}^N \mid \mathbf{C}\mathbf{n} \leq \mathbf{b}\} \quad \text{with} \quad \mathbf{C} \in \mathbb{R}^{k \times N}, \mathbf{b} \in \mathbb{R}^k,$$

we obtain the posterior distribution

$$p_{\text{posterior}}(\mathbf{n}) \propto \exp(-\frac{1}{2}\|\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}^{-\frac{1}{2}}(\mathbf{K}_N \mathbf{n} - \mathbf{e})\|_2^2 - \frac{1}{2}\gamma\|\mathbf{n}\|_2^2)I_S(\mathbf{n}). \quad (2.4)$$

The quantity of interest  $\mathbf{n}$  is estimated by computing the maximizer of the posterior distribution which is called the *maximum a posteriori estimator* (MAP). It is obtained by solving the quadratic programming problem

$$\mathbf{n}_{\text{MAP}}^\gamma := \underset{\mathbf{n} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}^{-\frac{1}{2}}(\mathbf{K}_N \mathbf{n} - \mathbf{e})\|_2^2 + \frac{1}{2}\gamma\|\mathbf{n}\|_2^2 \quad \text{s.t.} \quad \mathbf{C}\mathbf{n} \leq \mathbf{b}. \quad (2.5)$$

Note that  $\|\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}^{-\frac{1}{2}}(\mathbf{K}_N \mathbf{n} - \mathbf{e})\|_2^2 \sim \chi^2(N_l)$ , which gives  $\mathbb{E}(\|\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}^{-\frac{1}{2}}(\mathbf{K}_N \mathbf{n} - \mathbf{e})\|_2^2) = N_l$ .

A classical residual-based inference method is the so-called *discrepancy principle*. After selecting a *Morozov safety factor*  $\tau$  the regularization parameter  $\gamma$  is determined by demanding  $\|\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}^{-\frac{1}{2}}(\mathbf{K}_N \mathbf{n} - \mathbf{e})\|_2^2 = \tau N_l$ . A common choice for the safety factor is  $\tau = 1.1$ . We will give a more thorough introduction to the discrepancy principle and some results on it in Section 3.

*Monte Carlo methods* offer another way to evaluate the posterior distribution, cf. [6]. The advantage of Monte Carlo methods is that they take more of the statistical behavior of the observed measurement noise into account because all possible solutions with nonnegligible posterior probability are sampled and contribute to the inference result. However these methods require a lot of computational resources, which we cannot afford because our application requires that one FASP measurement data inversion must be completed in under thirty seconds using a regular notebook.

The discrepancy principle gets along with much less computational effort, but it does not take into account the specific shape of the distribution of the observed measurement noise. It does not explore the posterior distribution thoroughly and might give unreasonable results because of this.

In our hybrid approach we combine the advantages of both methods. We review Tikhonov regularization under linear constraints and derive conditions for the existence of a bijection between the regularization parameter and the residual. If these conditions are fulfilled, we can propose a set of regularization parameters obtained with the discrepancy principle using a set of Morozov safety factors corresponding to high-probability values of the weighted norm of the residual,  $\|\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}^{-\frac{1}{2}}(\mathbf{K}_N \mathbf{n} - \mathbf{e})\|_2^2 \sim \chi^2(N_l)$ . After this a Bayesian model-comparison procedure is applied to these reconstructions, and we rank them according to their posterior probabilities.

We show with numerical simulations that our method satisfies the demands on runtime and accuracy and that it is superior to existing inversion methods based on classical model-selection approaches. In the last section we extend our method to investigate two-component aerosols.

### 3 Tikhonov Regularization under Linear Constraints

Computing the maximum a posteriori estimator leads to a quadratic programming problem of the form

$$\mathbf{n}_\gamma := \underset{\mathbf{n} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{K}\mathbf{n} - \mathbf{r}\|_2^2 + \frac{1}{2} \gamma \|\mathbf{n}\|_2^2 \quad \text{s.t.} \quad \mathbf{C}\mathbf{n} \leq \mathbf{b}, \quad (3.1)$$

with  $\mathbf{K} := \Sigma_\sigma^{-\frac{1}{2}} \mathbf{K}_N$  and  $\mathbf{r} := \Sigma_\sigma^{-\frac{1}{2}} \mathbf{e}$ . The function to be minimized is known as the Tikhonov functional.

It is proved in [7] that the residual of the Tikhonov-regularized solution under linear constraints decreases monotonically with the regularization parameter  $\gamma$ . To the best of our knowledge conditions for *strict* monotonicity have not been found yet, so we derive some in the following. The advantage of having a strictly monotonic relation between regularization parameter and residual is that it gives a bijection. Thus we can then identify any regularization parameter  $\gamma$  from the range  $[0, \infty)$  with a unique residual value  $\|\mathbf{K}\mathbf{n}_\gamma - \mathbf{r}\|_2^2$  from the range  $[\|\mathbf{K}\mathbf{n}_0 - \mathbf{r}\|_2^2, \|\mathbf{K}\mathbf{n}_\infty - \mathbf{r}\|_2^2)$ . Here

$$\mathbf{n}_\infty := \underset{\mathbf{n} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{n}\|_2^2 \quad \text{s.t.} \quad \mathbf{C}\mathbf{n} \leq \mathbf{b} \quad (3.2)$$

is the *minimum norm element*. As shown in [8] there holds  $\lim_{\gamma \rightarrow \infty} \mathbf{n}_\gamma = \mathbf{n}_\infty$ . When our monotonicity conditions are satisfied, we obtain a set of distinct regularization parameters by proposing a set of distinct residual values from the range  $[\|\mathbf{K}\mathbf{n}_0 - \mathbf{r}\|_2^2, \|\mathbf{K}\mathbf{n}_\infty - \mathbf{r}\|_2^2)$ . The disadvantageous case of multiple prior distributions corresponding to the same residual value can therefore not occur. Note that in practice the cases  $\gamma = 0$  and  $\gamma = \infty$  are inadmissible, since then the Tikhonov prior distribution is improper or degenerates to a point mass, so we always restrict ourselves to a finite range  $(0, \gamma_{\max}]$  with  $\gamma_{\max} < \infty$ .

#### 3.1 Necessary Conditions for Strict Monotonicity

The following theorem shows that  $\mathbf{n}_\alpha \neq \mathbf{n}_\beta$  for all  $\alpha > \beta$  is the only necessary condition needed for strict monotonicity.

**Lemma 3.1.** *Let  $\alpha > \beta \geq 0$  be arbitrary and  $\mathbf{n}_\alpha$  and  $\mathbf{n}_\beta$  the solutions of (3.1) for  $\gamma = \alpha$  and  $\gamma = \beta$  respectively. If there holds  $\mathbf{n}_\alpha \neq \mathbf{n}_\beta$  for all  $\alpha > \beta$ , then the residual  $\|\mathbf{K}\mathbf{n}_\gamma - \mathbf{r}\|_2^2$  is strictly increasing for growing  $\gamma$ .*

*Proof.* From the first-order necessary Karush-Kuhn-Tucker conditions for the problem (3.1) we have that for each  $\gamma$  there exists a vector  $\mathbf{q}_\gamma \in \mathbb{R}^k$  with

$$\mathbf{K}^T \mathbf{K} \mathbf{n}_\gamma - \mathbf{K}^T \mathbf{r} + \gamma \mathbf{n}_\gamma + \mathbf{C}^T \mathbf{q}_\gamma = 0 \quad (3.3)$$

$$\mathbf{C} \mathbf{n}_\gamma \leq \mathbf{b} \quad (3.4)$$

$$\operatorname{diag}(\mathbf{q}_\gamma) (\mathbf{C} \mathbf{n}_\gamma - \mathbf{b}) = 0 \quad (3.5)$$

$$\mathbf{q}_\gamma \geq 0. \quad (3.6)$$

We define the difference vector

$$\mathbf{x} := \mathbf{n}_\beta - \mathbf{n}_\alpha$$

and subtract (3.3) for  $\gamma = \alpha$  with the same equation for  $\gamma = \beta$  to get

$$\mathbf{K}^T \mathbf{K} \mathbf{x} + \beta \mathbf{n}_\beta - \alpha \mathbf{n}_\alpha + \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) = 0. \quad (3.7)$$

Taking the scalar product of (3.7) with  $\mathbf{n}_\alpha$  and then with  $\mathbf{n}_\beta$  gives

$$\begin{aligned} \langle \mathbf{n}_\alpha, \mathbf{K}^T \mathbf{K} \mathbf{x} + \beta \mathbf{n}_\beta - \alpha \mathbf{n}_\alpha + \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle &= 0 \\ \text{and } \langle \mathbf{n}_\beta, \mathbf{K}^T \mathbf{K} \mathbf{x} + \beta \mathbf{n}_\beta - \alpha \mathbf{n}_\alpha + \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle &= 0. \end{aligned}$$

Our next step is to add  $(\alpha - \beta) \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle$  on both sides of the first relation and analogously  $(\alpha - \beta) \langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle$  on both sides of the latter relation, which results in

$$\begin{aligned} \langle \mathbf{n}_\alpha, (\mathbf{K}^T \mathbf{K} + \beta \mathbf{I}) \mathbf{x} + \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle &= (\alpha - \beta) \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle \\ \text{and} \quad \langle \mathbf{n}_\beta, (\mathbf{K}^T \mathbf{K} + \alpha \mathbf{I}) \mathbf{x} + \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle &= (\alpha - \beta) \langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle. \end{aligned}$$

Taking the difference of these two equations gives

$$\begin{aligned} &(\alpha - \beta) (\langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle - \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle) \\ &= \langle \mathbf{x}, (\mathbf{K}^T \mathbf{K} + \alpha \mathbf{I}) \mathbf{n}_\beta \rangle - \langle \mathbf{x}, (\mathbf{K}^T \mathbf{K} + \beta \mathbf{I}) \mathbf{n}_\alpha \rangle + \langle \mathbf{x}, \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle. \end{aligned}$$

On the one hand this implies

$$\begin{aligned} &(\alpha - \beta) (\langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle - \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle) \\ &= \langle \mathbf{x}, (\mathbf{K}^T \mathbf{K} + \beta \mathbf{I}) \mathbf{n}_\beta \rangle + (\alpha - \beta) \langle \mathbf{x}, \mathbf{n}_\beta \rangle - \langle \mathbf{x}, (\mathbf{K}^T \mathbf{K} + \beta \mathbf{I}) \mathbf{n}_\alpha \rangle + \langle \mathbf{x}, \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle \\ &= \langle \mathbf{x}, (\mathbf{K}^T \mathbf{K} + \beta \mathbf{I}) \mathbf{x} \rangle + (\alpha - \beta) \langle \mathbf{x}, \mathbf{n}_\beta \rangle + \langle \mathbf{x}, \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle, \end{aligned}$$

while on the other hand

$$\begin{aligned} &(\alpha - \beta) (\langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle - \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle) \\ &= \langle \mathbf{x}, (\mathbf{K}^T \mathbf{K} + \alpha \mathbf{I}) \mathbf{n}_\beta \rangle - \langle \mathbf{x}, (\mathbf{K}^T \mathbf{K} + \alpha \mathbf{I}) \mathbf{n}_\alpha \rangle - (\beta - \alpha) \langle \mathbf{x}, \mathbf{n}_\alpha \rangle + \langle \mathbf{x}, \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle \\ &= \langle \mathbf{x}, (\mathbf{K}^T \mathbf{K} + \alpha \mathbf{I}) \mathbf{x} \rangle + (\alpha - \beta) \langle \mathbf{x}, \mathbf{n}_\alpha \rangle + \langle \mathbf{x}, \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle \end{aligned}$$

holds. Adding these gives

$$\begin{aligned} &2(\alpha - \beta) (\langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle - \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle) \\ &= \langle \mathbf{x}, (2\mathbf{K}^T \mathbf{K} + (\alpha + \beta)\mathbf{I}) \mathbf{x} \rangle + (\alpha - \beta) (\langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle - \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle) + 2 \langle \mathbf{x}, \mathbf{C}^T (\mathbf{q}_\beta - \mathbf{q}_\alpha) \rangle, \end{aligned}$$

and finally we arrive at

$$\begin{aligned} &(\alpha - \beta) (\langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle - \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle) \\ &= \langle \mathbf{n}_\beta - \mathbf{n}_\alpha, (2\mathbf{K}^T \mathbf{K} + (\alpha + \beta)\mathbf{I}) (\mathbf{n}_\beta - \mathbf{n}_\alpha) \rangle + 2 \langle \mathbf{C} (\mathbf{n}_\beta - \mathbf{n}_\alpha), \mathbf{q}_\beta - \mathbf{q}_\alpha \rangle. \end{aligned} \tag{3.8}$$

Now we consider the term  $\langle \mathbf{C} (\mathbf{n}_\beta - \mathbf{n}_\alpha), \mathbf{q}_\beta - \mathbf{q}_\alpha \rangle$ . The following four cases can occur:

$i$ -th constraint for the Tikhonov functional for		$(\mathbf{C}\mathbf{n}_\beta)_i$	$(\mathbf{C}\mathbf{n}_\alpha)_i$	$(\mathbf{C}\mathbf{n}_\beta)_i$ $-(\mathbf{C}\mathbf{n}_\alpha)_i$	$(\mathbf{q}_\beta)_i$	$(\mathbf{q}_\alpha)_i$	$(\mathbf{q}_\beta)_i$ $-(\mathbf{q}_\alpha)_i$
$\gamma = \beta$	$\gamma = \alpha$						
active	active	$= (\mathbf{b})_i$	$= (\mathbf{b})_i$	$= 0$	$\geq 0$	$\geq 0$	void
inactive	active	$< (\mathbf{b})_i$	$= (\mathbf{b})_i$	$< 0$	$= 0$	$\geq 0$	$\leq 0$
active	inactive	$= (\mathbf{b})_i$	$< (\mathbf{b})_i$	$> 0$	$\geq 0$	$= 0$	$\geq 0$
inactive	inactive	$< (\mathbf{b})_i$	$< (\mathbf{b})_i$	void	$= 0$	$= 0$	$= 0$

From this we see that all components of the vector

$$\text{diag}(\mathbf{C} (\mathbf{n}_\beta - \mathbf{n}_\alpha)) (\mathbf{q}_\beta - \mathbf{q}_\alpha)$$

are nonnegative, and so  $\langle \mathbf{C} (\mathbf{n}_\beta - \mathbf{n}_\alpha), \mathbf{q}_\beta - \mathbf{q}_\alpha \rangle \geq 0$ . Under the assumption  $\mathbf{n}_\alpha \neq \mathbf{n}_\beta$  we have  $\mathbf{x} \neq 0$ , and since the matrix  $2\mathbf{K}^T \mathbf{K} + (\alpha + \beta)\mathbf{I}$  is positive definite we finally conclude with (3.8) that

$$(\alpha - \beta) (\langle \mathbf{n}_\beta, \mathbf{n}_\beta \rangle - \langle \mathbf{n}_\alpha, \mathbf{n}_\alpha \rangle) > 0,$$

which is equivalent to  $\|\mathbf{n}_\beta\|_2^2 > \|\mathbf{n}_\alpha\|_2^2$ .

We proceed then with

$$\|\mathbf{K}\mathbf{n}_\alpha - \mathbf{r}\|_2^2 - \|\mathbf{K}\mathbf{n}_\beta - \mathbf{r}\|_2^2 = \langle \mathbf{x}, \mathbf{K}^T \mathbf{K} \mathbf{x} \rangle + 2 \langle -\mathbf{x}, \mathbf{K}^T \mathbf{K} \mathbf{n}_\beta - \mathbf{K}^T \mathbf{r} + \beta \mathbf{n}_\beta \rangle + 2\beta \langle \mathbf{x}, \mathbf{n}_\beta \rangle \quad (3.9)$$

by using  $\mathbf{n}_\alpha = \mathbf{n}_\beta - \mathbf{x}$ . The variational inequality for the Tikhonov functional for  $\gamma = \beta$  yields  $\langle -\mathbf{x}, \mathbf{K}^T \mathbf{K} \mathbf{n}_\beta - \mathbf{K}^T \mathbf{r} + \beta \mathbf{n}_\beta \rangle \geq 0$ . Moreover we have

$$\langle \mathbf{x}, \mathbf{n}_\beta \rangle = \langle \mathbf{n}_\beta - \mathbf{n}_\alpha, \mathbf{n}_\beta \rangle \geq \|\mathbf{n}_\beta\|_2^2 - \|\mathbf{n}_\alpha\|_2 \|\mathbf{n}_\beta\|_2 > 0.$$

In summary we have shown  $\|\mathbf{K}\mathbf{n}_\alpha - \mathbf{r}\|_2^2 > \|\mathbf{K}\mathbf{n}_\beta - \mathbf{r}\|_2^2$ .  $\square$

**Remark 3.2.** From (3.9) follows that all  $\mathbf{n}_\gamma$  with  $\|\mathbf{K}\mathbf{n}_\gamma - \mathbf{r}\|_2 = \tau$  for an arbitrary but fixed  $\tau$  must coincide. This means in other words that if the residual of the regularized solutions ‘gets stuck’ at some value  $\tau$ , the solutions  $\mathbf{n}_\gamma$  are constant for these values of  $\gamma$ . In the next section we derive conditions which prevent this case.

### 3.2 Sufficient Conditions for Strict Monotonicity

In this section we derive sufficient conditions for  $\mathbf{n}_\alpha \neq \mathbf{n}_\beta$  for  $\alpha > \beta$ , hence by Lemma 3.1 for strict monotonicity. In particular we focus on constraints of the form  $\mathbf{C}\mathbf{n} \geq 0$  with  $\mathbf{C} \in \mathbb{R}^{k \times N}$  and  $k \leq N$ , i.e. on generalized nonnegativity constraints. For this specific type of constraints we have that for the minimum norm solution  $\mathbf{n}_\infty$  defined in (3.2) that  $\mathbf{n}_\infty \equiv 0$  holds, which gives according to [8] the relation  $\|\mathbf{K}\mathbf{n}_\alpha - \mathbf{r}\|_2 \leq \|\mathbf{r}\|_2$  for all  $\alpha \geq 0$ . We carry out our following considerations under following important assumption.

**Assumption 3.3.** *The regularization parameter  $\alpha$  is selected in such a way that for the regularized residual the inequality  $\|\mathbf{K}\mathbf{n}_\alpha - \mathbf{r}\|_2 \leq c\delta$  holds, where  $c > 0$  is a fixed constant and  $\delta > 0$  the noise level so that the relation  $c\delta < \|\mathbf{r}\|_2$  is satisfied.*

Under this assumption we have the strict inequality  $\|\mathbf{K}\mathbf{n}_\alpha - \mathbf{r}\|_2 < \|\mathbf{r}\|_2$  for all  $\alpha \in [0, \infty)$ .

**Remark 3.4.** If we assume  $\mathbf{K}\mathbf{n}_\alpha = \mathbf{r}_{true}$ , where  $\mathbf{r}_{true}$  is the ‘true’ data vector, we can rewrite the first part of above assumption as  $\|\mathbf{r} - \mathbf{r}_{true}\|_2 < c\delta$ . This is a standard assumption made in inverse problems literature, cf. [8].

**Theorem 3.5.** *Let  $\mathbf{n}_\alpha$  be given by*

$$\mathbf{n}_\alpha := \underset{\mathbf{n} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{K}\mathbf{n} - \mathbf{r}\|_2^2 + \frac{1}{2} \alpha \|\mathbf{n}\|_2^2 \quad \text{s.t.} \quad -\mathbf{C}\mathbf{n} \leq 0, \quad (3.10)$$

with  $\mathbf{C} \in \mathbb{R}^{k \times N}$  having full row rank  $k \leq N$ . If  $\|\mathbf{K}\mathbf{n}_\alpha - \mathbf{r}\|_2 < \|\mathbf{r}\|_2$ , or equivalently  $\mathbf{n}_\alpha \neq 0$  for all  $\alpha \in [0, \infty)$  according to Lemma 3.1 and Remark 3.2, then we have  $\mathbf{n}_\alpha \neq \mathbf{n}_\beta$  for all  $\alpha > \beta$ .

*Proof.* Let  $\alpha > \beta$ . Let  $\mathbf{C}_{act}^\alpha$  denote the submatrix of  $\mathbf{C}$  with active constraints in (3.10) for the regularization parameter  $\alpha$ .

We first consider the case  $\mathbf{C}_{act}^\alpha \neq \mathbf{C}_{act}^\beta$ . We obtain  $\mathbf{C}(\mathbf{n}_\alpha - \mathbf{n}_\beta) \neq 0$ , i.e.  $\mathbf{n}_\alpha - \mathbf{n}_\beta \notin \ker(\mathbf{C})$ . This gives directly  $\mathbf{n}_\alpha - \mathbf{n}_\beta \neq 0$ .

Now we turn to the case  $\mathbf{C}_{act}^\alpha = \mathbf{C}_{act}^\beta$ . The first-order necessary conditions for a minimizer in (3.10) are given by

$$\mathbf{K}^T \mathbf{K} \mathbf{n} - \mathbf{K}^T \mathbf{r} + \alpha \mathbf{n} - \mathbf{C}^T \mathbf{q}_\alpha = 0, \quad (3.11)$$

where  $\mathbf{q}_\alpha \geq 0$ . Let us assume  $\mathbf{n}_\alpha = \mathbf{n}_\beta$ . Then taking the difference of (3.11) for the parameters  $\alpha$  and  $\beta$  yields

$$(\alpha - \beta) \mathbf{n}_\alpha - \mathbf{C}^T (\mathbf{q}_\alpha - \mathbf{q}_\beta) = 0. \quad (3.12)$$

Let us first consider the subcase that none of the constraints is active. Then we have  $\mathbf{q}_\alpha = \mathbf{q}_\beta = 0$ , which implies  $(\alpha - \beta) \mathbf{n}_\alpha = 0$ . This contradicts  $\mathbf{n}_\alpha \neq 0$ , so we must have  $\mathbf{n}_\alpha \neq \mathbf{n}_\beta$ . Now we turn to the subcase that at least one constraint is active. Let  $\mathbf{q}_{act}^\alpha$  and  $\mathbf{q}_{act}^\beta$  be the subvectors of  $\mathbf{q}_\alpha$  and  $\mathbf{q}_\beta$  corresponding to active constraints. Then we can rewrite the last equation as

$$(\alpha - \beta) \mathbf{n}_\alpha - \mathbf{C}_{act}^{\alpha T} (\mathbf{q}_{act}^\alpha - \mathbf{q}_{act}^\beta) = 0,$$

where we remember that  $\mathbf{C}_{act}^\alpha$  is obtained from  $\mathbf{C}$  by canceling its  $i$ -th row when the constraint  $-(\mathbf{C}\mathbf{n})_i \leq 0$  is inactive and thus  $(\mathbf{q}_\alpha)_i = (\mathbf{q}_\beta)_i = 0$  holds. Our next step is to multiply this equation from the left with  $\mathbf{C}_{act}^\alpha$ . By construction of  $\mathbf{C}_{act}^\alpha$  we have  $\mathbf{C}_{act}^\alpha \mathbf{n}_\alpha = 0$  and therefore

$$-\mathbf{C}_{act}^\alpha \mathbf{C}_{act}^{\alpha T} (\mathbf{q}_{act}^\alpha - \mathbf{q}_{act}^\beta) = 0.$$

Since  $\mathbf{C}_{act}^\alpha \mathbf{C}_{act}^{\alpha T}$  has full rank, this implies  $\mathbf{q}_{act}^\alpha = \mathbf{q}_{act}^\beta$  and hence  $\mathbf{q}_\alpha = \mathbf{q}_\beta$ . Inserting this finding back into (3.12) gives  $\mathbf{n}_\alpha = 0$ , which contradicts our assumption  $\mathbf{n}_\alpha \neq 0$  for all  $\alpha \in [0, \infty)$ . Therefore we must also have  $\mathbf{n}_\alpha \neq \mathbf{n}_\beta$  in this subcase.  $\square$

### 3.3 The Discrepancy Principle

With the next Theorem we summarize our previous results.

**Theorem 3.6.** *Let the conditions of Theorem 3.5 be fulfilled and let  $\mathbf{n}_\infty$  be the minimum norm solution defined in (3.2). Define  $r_0 := \|\mathbf{K}\mathbf{n}_0 - \mathbf{r}\|_2$  and  $r_\infty := \|\mathbf{K}\mathbf{n}_\infty - \mathbf{r}\|_2$ . Then there exist for any  $\tau$  from  $[r_0, r_\infty)$  a unique  $\gamma$  from  $[0, \infty)$  such that  $\|\mathbf{K}\mathbf{n}_\gamma - \mathbf{r}\|_2 = \tau$ . The residual grows strictly monotonically with  $\gamma$ .*

$\square$

**Remark 3.7.** The discrepancy principle carries directly over to generalized Tikhonov regularization, where the prior distribution is given by

$$p_{prior}(\mathbf{n}) \propto \exp(-\frac{1}{2}\gamma \mathbf{n}^T \mathbf{R} \mathbf{n}) I_S(\mathbf{n}),$$

where  $\mathbf{R}$  is a positive definite regularization matrix and  $I_S(\mathbf{n})$  is the indicator function of  $S = \{\mathbf{n} \in \mathbb{R}^N \mid -\mathbf{C}\mathbf{n} \leq 0\}$ . Here we have to solve the quadratic programming problem

$$\min_{\mathbf{n} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{K}\mathbf{n} - \mathbf{r}\|_2^2 + \frac{1}{2} \gamma \mathbf{n}^T \mathbf{R} \mathbf{n} \quad \text{s.t.} \quad -\mathbf{C}\mathbf{n} \leq 0.$$

Let  $\mathbf{R} = \mathbf{U}^T \mathbf{U}$  the Cholesky decomposition. Then the substitution  $\mathbf{n} = \mathbf{U}^{-1} \mathbf{v}$  transforms the above quadratic programming problem into the standard form (3.1).

### 3.4 Convergence Analysis

At this point we review some classical convergence criteria for parameter-choice strategies for Tikhonov regularization under linear constraints. With convergence we mean that the regularized reconstructions approach the true solution of the noise-free linear inverse problem as the noise level goes to 0. We decompose the noisy data vector  $\mathbf{r}$  into

$$\mathbf{r} = \Sigma_\sigma^{-\frac{1}{2}} (\mathbf{e}_{true} + \boldsymbol{\delta})$$

with  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{N_l})^T, \quad \delta_i \sim \mathcal{N}(0, \sigma_i^2).$

We carry out our convergence analysis under following assumption.

**Assumption 3.8.** *The covariance matrix  $\Sigma_\sigma$  has the simple form*

$$\Sigma_\sigma = \delta^2 \cdot \text{diag}(\sigma_1^2, \dots, \sigma_{N_l}^2) =: \delta^2 \cdot \Sigma,$$

where  $\delta \geq 0$  is an arbitrary but fixed noise level and  $\sigma_1, \dots, \sigma_{N_l}$  are fixed.

Now instead of maximizing the posterior probability (2.4) directly, we use the fact that

$$\exp(-\frac{1}{2} \|\Sigma^{-\frac{1}{2}} (\mathbf{K}_N \mathbf{n} - (\mathbf{e}_{true} + \boldsymbol{\delta}))\|_2^2 - \frac{1}{2} \gamma \delta^2 \|\mathbf{n}\|_2^2) I_S(\mathbf{n})$$

has the same maximizer. To obtain the function above we scaled the argument of the exponential in (2.4) with the noise level  $\delta^2$ . For simpler notation we redefine for all the following

$$\mathbf{K} := \Sigma^{-\frac{1}{2}} \mathbf{K}_N, \quad \mathbf{r} := \Sigma^{-\frac{1}{2}} (\mathbf{e}_{true} + \boldsymbol{\delta}) \quad \text{and} \quad \alpha = \gamma \delta^2.$$

This means that we work with versions of  $\Sigma_{\sigma}^{-\frac{1}{2}}(\mathbf{e}_{true} + \delta)$  and  $\Sigma_{\sigma}^{-\frac{1}{2}}\mathbf{K}_N$  where the noise magnitude  $\delta^2$  is scaled out. So instead of solving (3.1), we now solve

$$\min_{\mathbf{n} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{K}\mathbf{n} - \mathbf{r}\|_2^2 + \frac{1}{2}\alpha \|\mathbf{n}\|_2^2 \quad \text{s.t.} \quad \mathbf{C}\mathbf{n} \leq \mathbf{b}. \quad (3.13)$$

We have to point out that the back-scaled parameter  $\gamma = \alpha/\delta^2$  must be used for the statistical computations for the posterior probabilities in Section 3.5. So we always compute the parameter  $\alpha$  first from (3.13) and then obtain  $\gamma$  from it. We can already see here that Bayesian model-selection computations are not feasible for very small noise levels  $\delta$ , since the parameter  $\gamma$  diverges as  $\delta$  tends to 0. Another reason for skipping the model selection step for  $\delta$  approaching 0 is that the entries of the covariance matrix  $\Sigma_{\sigma}$  get closer to 0 as well here, which causes problems in the statistical computations which will follow in Section 3.5. We recommend to switch to the classical discrepancy principle in this case.

Now we present the standard convergence rate for Tikhonov regularization.

**Proposition 3.9.** *If the noise-free true solution  $\mathbf{n}_0$  is an element of the feasible set of (3.1), then the regularized solutions  $\mathbf{n}_{\alpha}$  of the noise-free problem satisfies*

$$\|\mathbf{K}(\mathbf{n}_0 - \mathbf{n}_{\alpha})\|_2 = \mathcal{O}(\alpha^{\frac{1}{2}}) \quad (3.14)$$

as  $\alpha$  goes to 0. Thus  $\lim_{\alpha \rightarrow 0} \mathbf{n}_{\alpha} = \mathbf{n}_0$ .

*Proof.* Rearranging (3.8) for  $\beta = 0$  gives the error representation

$$\|\mathbf{K}(\mathbf{n}_0 - \mathbf{n}_{\alpha})\|_2^2 = \langle \mathbf{n}_0 - \mathbf{n}_{\alpha}, \alpha \mathbf{n}_{\alpha} - \mathbf{C}^T(\mathbf{q}_0 - \mathbf{q}_{\alpha}) \rangle.$$

Now since  $\langle \mathbf{n}_0 - \mathbf{n}_{\alpha}, \mathbf{C}^T(\mathbf{q}_0 - \mathbf{q}_{\alpha}) \rangle \geq 0$  and  $\|\mathbf{n}_{\alpha}\|_2 \leq \|\mathbf{n}_0\|_2$  hold, we can therefore estimate

$$\|\mathbf{K}(\mathbf{n}_0 - \mathbf{n}_{\alpha})\|_2^2 \leq \alpha \|\mathbf{n}_0\|_2^2$$

which gives the first result. The second assertion was proved in [8].  $\square$

**Proposition 3.10.** *Let  $\mathbf{r}$  and  $\tilde{\mathbf{r}}$  be two different data vectors for (3.13) and let  $\mathbf{n}_{\alpha}$  and  $\tilde{\mathbf{n}}_{\alpha}$  be the corresponding regularized solutions of (3.13) for the parameter  $\alpha$ . Then*

$$\|\mathbf{K}(\mathbf{n}_{\alpha} - \tilde{\mathbf{n}}_{\alpha})\|_2 \leq \|\mathbf{r} - \tilde{\mathbf{r}}\|_2 \quad \text{and} \quad \|\mathbf{n}_{\alpha} - \tilde{\mathbf{n}}_{\alpha}\|_2 \leq \frac{\|\mathbf{r} - \tilde{\mathbf{r}}\|_2}{\alpha^{\frac{1}{2}}}. \quad (3.15)$$

*Proof.* We give the proof from [8]. The solutions  $\mathbf{n}_{\alpha}$  and  $\tilde{\mathbf{n}}_{\alpha}$  fulfill the variational inequalities

$$\begin{aligned} \langle \mathbf{K}^T \mathbf{K} \mathbf{n}_{\alpha} - \mathbf{K}^T \mathbf{r} + \alpha \mathbf{n}_{\alpha}, \tilde{\mathbf{n}}_{\alpha} - \mathbf{n}_{\alpha} \rangle &\geq 0 \\ \text{and} \quad \langle \mathbf{K}^T \mathbf{K} \tilde{\mathbf{n}}_{\alpha} - \mathbf{K}^T \tilde{\mathbf{r}} + \alpha \tilde{\mathbf{n}}_{\alpha}, \mathbf{n}_{\alpha} - \tilde{\mathbf{n}}_{\alpha} \rangle &\geq 0. \end{aligned}$$

Adding them gives

$$\begin{aligned} \|\mathbf{K}(\tilde{\mathbf{n}}_{\alpha} - \mathbf{n}_{\alpha})\|_2^2 + \alpha \|\tilde{\mathbf{n}}_{\alpha} - \mathbf{n}_{\alpha}\|_2^2 &\leq \langle \tilde{\mathbf{r}} - \mathbf{r}, \mathbf{K}(\tilde{\mathbf{n}}_{\alpha} - \mathbf{n}_{\alpha}) \rangle \\ &\leq \|\tilde{\mathbf{r}} - \mathbf{r}\|_2 \|\mathbf{K}(\tilde{\mathbf{n}}_{\alpha} - \mathbf{n}_{\alpha})\|_2, \end{aligned}$$

and the desired results follow from the last inequality.  $\square$

Finally we show under which conditions the regularized solutions  $\mathbf{n}_{\alpha}^{\delta}$  of the noisy problem (3.13) converge to the true solution  $\mathbf{n}_0$  of the noise-free problem for  $\delta \rightarrow 0$ . In preparation we note that for the weighted residual with noise level  $\delta$

$$\begin{aligned} \|\Sigma_{\sigma}^{-\frac{1}{2}}\delta\|_2^2 &\sim \chi^2(N_l) \quad \text{thus} \quad \mathbb{E}(\|\Sigma_{\sigma}^{-\frac{1}{2}}\delta\|_2^2) = \delta^2 \cdot \mathbb{E}(\|\Sigma_{\sigma}^{-\frac{1}{2}}\delta\|_2^2) = N_l \delta^2 = \mathcal{O}(\delta^2), \\ \text{i.e.} \quad \mathbb{E}(\|\Sigma_{\sigma}^{-\frac{1}{2}}\delta\|_2) &\leq \left( \mathbb{E}(\|\Sigma_{\sigma}^{-\frac{1}{2}}\delta\|_2^2) \right)^{\frac{1}{2}} = \mathcal{O}(\delta). \end{aligned}$$

We set  $\mathbf{r}_{true} := \Sigma^{-\frac{1}{2}}\mathbf{e}_{true}$ . Then for the expected value we have

$$\mathbb{E}(\|\mathbf{r} - \mathbf{r}_{true}\|_2) = \mathbb{E}(\|\Sigma^{-\frac{1}{2}}\delta\|_2) = \mathcal{O}(\delta).$$



**Theorem 3.11.** *If we have  $\mathbb{E}(\|\mathbf{r} - \mathbf{r}_{true}\|_2) = \mathcal{O}(\delta)$  and  $\alpha(\delta)$  has the properties  $\lim_{\delta \rightarrow 0} \alpha(\delta) = 0$  and  $\lim_{\delta \rightarrow 0} \frac{\delta^2}{\alpha(\delta)} = 0$ , then  $\lim_{\delta \rightarrow 0} \mathbb{E}(\|\mathbf{n}_{\alpha(\delta)}^\delta - \mathbf{n}_0\|_2) = 0$  holds.*

*Proof.* We have

$$\mathbb{E}(\|\mathbf{n}_{\alpha(\delta)}^\delta - \mathbf{n}_0\|_2) \leq \mathbb{E}(\|\mathbf{n}_{\alpha(\delta)}^\delta - \mathbf{n}_{\alpha(\delta)}\|_2) + \mathbb{E}(\|\mathbf{n}_{\alpha(\delta)} - \mathbf{n}_0\|_2),$$

where  $\mathbf{n}_{\alpha(\delta)}$  is the regularized solution for the noise-free data  $\mathbf{r}_{true}$ . Having  $\mathbb{E}(\|\mathbf{r} - \mathbf{r}_{true}\|_2) = \mathcal{O}(\delta)$ , we can further estimate using Proposition 3.10

$$\mathbb{E}(\|\mathbf{n}_{\alpha(\delta)}^\delta - \mathbf{n}_0\|_2) \leq \mathbb{E}(\|\mathbf{n}_{\alpha(\delta)} - \mathbf{n}_0\|_2) + \frac{\mathcal{O}(\delta)}{\alpha^{\frac{1}{2}}(\delta)}.$$

Then the result follows with Proposition 3.9.  $\square$

### 3.5 Model Generation under Nonnegativity Constraints

Suppose we have discretized our linear operator with a Galerkin collocation method on a set of  $m$  different grids. Each grid has  $N_k$  collocation points with  $N_1 < \dots < N_m$  and we have computed a discrete approximation  $\mathbf{K}_k$  to  $K$  for each grid. The approximation  $\mathbf{n}_k$  of the sought-after function  $n$  lies in  $\mathbb{R}^{N_k}$ . For each grid we apply a Tikhonov prior with nonnegativity constraints on the observed model uncertainty such that we have according to the previously derived results a bijection between attainable residuals and regularization parameters.

Because  $\delta_1, \dots, \delta_{N_l}$  are normally distributed, it follows with  $\Sigma_\sigma^{-\frac{1}{2}} \mathbf{K}_k \mathbf{n}_k = \mathbf{e}_{true}$  that

$$\|\Sigma_\sigma^{-\frac{1}{2}} (\mathbf{K}_k \mathbf{n}_k - (\mathbf{e}_{true} + \boldsymbol{\delta}))\|_2^2 \sim \chi^2(N_l),$$

$$\text{and thus } \mathbb{E}(\|\Sigma_\sigma^{-\frac{1}{2}} (\mathbf{K}_k \mathbf{n}_k - (\mathbf{e}_{true} + \boldsymbol{\delta}))\|_2^2) = N_l, \quad \forall k \in \{1, \dots, m\}.$$

In the literature on the discrepancy principle, e.g. in [9], the error estimate  $N_l$  is multiplied with a factor  $\tau$  near 1 which is known as *Morozov's safety parameter*. Now we interpret it here statistically as high-probability values of the observed distribution of the weighted residual. Of course we do not select just one single value for  $\tau$ , instead we select a grid of Morozov safety parameters  $\tau_1, \dots, \tau_s$ . The following example of the  $\chi^2(48)$  probability density functions illustrates this strategy:

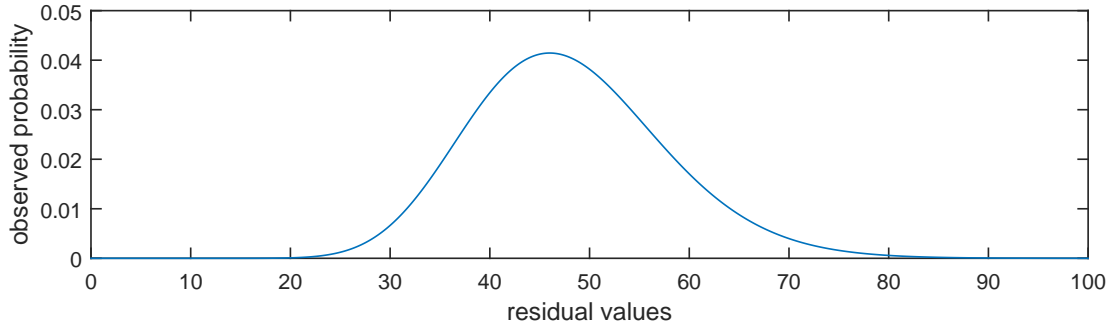


Figure 3.1:  $\chi^2(48)$  probability density function

It is indeed a unimodal distribution with residual values having a nonnegligible probability ranging from 30 to 70. For  $N_l = 48$  this corresponds to values of  $\tau$  ranging from ca. 0.6 to ca. 1.5. Therefore proposing just a single residual value for the discrepancy principle ( $1.1N_l$  would be a common choice) excludes many probable reconstructions corresponding to other residual values, such that the posterior probability exploration is limited. Moreover the danger of under- or overregularization would be high.

As in the previous section we use the normalized version  $\Sigma$  of the covariance matrix  $\Sigma_\sigma$ . This means that we try to fit the normalized residuals  $\|\Sigma^{-\frac{1}{2}} (\mathbf{K}_k \mathbf{n}_k - (\mathbf{e}_{true} + \boldsymbol{\delta}))\|_2^2$  to the values  $\tau N_l \delta^2$  where the values for  $\tau$  run through the grid of preselected Morozov safety factors. In practice the noise magnitude  $\delta^2$  is taken as the biggest measurement sample mean and  $\Sigma$  is estimated from  $\Sigma_\sigma$  by normalizing it with the estimate for  $\delta^2$ .

For the following we set  $\mathbf{e}_{real} := (\tilde{e}_1, \dots, \tilde{e}_{N_l})^T$ , hence this is the vector of the realizations of the random variables  $e_1, \dots, e_{N_l}$ . With these preparations the model generation step proceeds as follows:

---

**Algorithm 1** Model Generation

---

```
1:  $MaxDisc = 3$ 
2:  $SolutionSets = \{\}$ 
3:  $ApproxSets = \{\}$ 
4:  $PriorSets = \{\}$ 
5:  $TauSets = \{\}$ 
6:  $DiscCntr = 0$ 
7: estimate  $\sigma_1^2, \dots, \sigma_{N_l}^2$  from the sample means approximating the standard deviations of  $e_1, \dots, e_{N_l}$ .
8:  $\delta^2 := \max\{\sigma_1^2, \dots, \sigma_{N_l}^2\}$ 
9:  $\Sigma := \delta^{-2} \cdot \text{diag}(\sigma_1^2, \dots, \sigma_{N_l}^2)$ 
10: for  $i = 1$  to  $m$  do
11:    $S_i = \{\}$ 
12:    $A_i = \{\}$ 
13:    $P_i = \{\}$ 
14:    $T_i = \{\}$ 
15:    $\mathbf{n}_{lsqnnng} = \underset{\mathbf{n} \in \mathbb{R}^{N_i}}{\text{argmin}} \frac{1}{2} \|\Sigma^{-\frac{1}{2}}(\mathbf{K}_i \mathbf{n} - \mathbf{e}_{real})\|_2^2$  s.t.  $\mathbf{n} \geq 0$ 
16:    $R_{lsqnnng} = \|\Sigma^{-\frac{1}{2}}(\mathbf{K}_i \mathbf{n}_{lsqnnng} - \mathbf{e}_{real})\|_2^2$ 
17:   for  $j = 1$  to  $s$  do
18:     if  $R_{lsqnnng} < \tau_j N_l \delta^2 \wedge \tau_j N_l \delta^2 < \|\Sigma^{-\frac{1}{2}} \mathbf{e}_{real}\|_2^2$  then
19:       compute  $\gamma_{ij}$  such that
20:        $\mathbf{n}_{trial} = \underset{\mathbf{n} \in \mathbb{R}^{N_i}}{\text{argmin}} \frac{1}{2} \|\Sigma^{-\frac{1}{2}}(\mathbf{K}_i \mathbf{n} - \mathbf{e}_{real})\|_2^2 + \frac{1}{2} \gamma_{ij} \mathbf{n}^T \mathbf{R}_i \mathbf{n}$  s.t.  $\mathbf{n} \geq 0$ 
21:       with  $\|\Sigma^{-\frac{1}{2}}(\mathbf{K}_i \mathbf{n}_{trial} - \mathbf{e}_{real})\|_2^2 = \tau_j N_l \delta^2$ 
22:     end if
23:     if  $\mathbf{n}_{trial}$  exists then
24:        $S_i = S_i \cup \{\mathbf{n}_{trial}\}$ 
25:        $A_i = A_i \cup \{\mathbf{K}_i\}$ 
26:        $P_i = P_i \cup \{\gamma_{ij} \mathbf{R}_i\}$ 
27:        $T_i = T_i \cup \{\tau_j\}$ 
28:     end if
29:   end for
30:   if  $S_i, A_i, P_i$  and  $T_i$  not empty then
31:      $SolutionSets = SolutionSets \cup \{S_i\}$ 
32:      $ApproxSets = ApproxSets \cup \{A_i\}$ 
33:      $PriorSets = PriorSets \cup \{P_i\}$ 
34:      $TauSets = TauSets \cup \{T_i\}$ 
35:      $DiscCntr = DiscCntr + 1$ 
36:   end if
37:   if  $DiscCntr == MaxDisc$  then
38:     break
39:   end if
40: end for
```

---

The outer loop runs through the discretization levels beginning with the coarsest one. This approach is in accordance with the principle of *Occam's razor*, where among all possible explanations of a problem simpler ones are preferred over more complicated ones. Another motivation is *regularization by discretization*, which means that the approximate problems for the operator inversion are for coarser discretizations less ill-conditioned than for finer discretizations. But by using the discrepancy principle we ensure that the models selected are not too coarse by demanding that the model has to fit the data, which means that the residuals may not be too big.

For each  $i$ -th discretization level in the outer loop, the inner loop runs through the preselected grid

of Morozov safety factors, where for each factor  $\tau_j$  the computation of a regularized solution  $\mathbf{n}_{trial}$  with residual  $\tau_j N_l$  is attempted. In line 18 it is checked if the discrepancy principle is applicable. If it is possible to compute  $\mathbf{n}_{trial}$ , this reconstruction is stored in the container  $S_i$  and the approximation  $\mathbf{K}_i$  to  $K$  in  $A_i$ . The prior information given by the regularization parameter  $\gamma_{ij}$  and the regularization matrix  $\frac{1}{2}\mathbf{R}_i$  are stored in  $P_i$  and the residual parameter  $\tau_j$  in  $T_i$ . These matrices will be used to compute the Bayesian posterior probabilities for the model selection in the next section.

If in the current discretization level the containers with reconstructions, operator approximation matrices, prior informations and residual parameters are not empty, they are added to the containers *SolutionSets*, *ApproxSets*, *PriorSets* and *TauSets* respectively. Note that we have limited the maximal number of admissible discretization levels to three. On the one hand this is done to save computational effort, but on the other hand it turns out that the posterior probabilities get too similar and thus not clearly or reliably distinguishable when using too many finely discretized models.

## 4 Model Selection

### 4.1 Posterior Model Probabilities under Nonnegativity Constraints

In this section we apply the Bayesian model selection framework as introduced in [10]. Since we assume that the data is given by independent Gaussian random variables, the observed model uncertainty is a multivariate Gaussian distribution. For any of the approximations  $\mathbf{K}_k$  to the operator  $K$  with  $k \in \{1, \dots, m\}$  it is given by

$$p(\mathbf{e}|\mathbf{n}, N_k, \mathbf{K}_k) = (2\pi)^{-\frac{N_l}{2}} |\det(\boldsymbol{\Sigma}_\sigma)|^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{K}_k \mathbf{n} - \mathbf{e})^T \boldsymbol{\Sigma}_\sigma^{-1} (\mathbf{K}_k \mathbf{n} - \mathbf{e})). \quad (4.1)$$

Here the vector  $\mathbf{n} \in \mathbb{R}^{N_k}$  represents all possible reconstructions for the current discretization.

We know beforehand that our reconstruction must be nonnegative and that it is smooth. We put this prior knowledge into our reconstruction method by setting up the Bayesian conditional prior probability which is determined by

$$p(\mathbf{n}|N_k, \mathbf{K}_k, \mathbf{R}_k, \gamma_{kj}) = C_{kj}^{-1} \exp(-\frac{1}{2}\gamma_{kj} \mathbf{n}^T \mathbf{R}_k \mathbf{n}) I_{\geq 0}(\mathbf{n}), \quad (4.2)$$

where  $I_{\geq 0}(\mathbf{n})$  is the indicator function of the first quadrant of  $\mathbb{R}^{N_k}$ ,  $\mathbf{R}_k$  is the regularization matrix and  $\gamma_{kj}$  is the regularization parameter. All these quantities were computed and stored in the model generation procedure in the previous section.

If  $\mathbf{R}_k$  is regular and positive definite, the normalizing constant

$$C_{kj} = \int_{[0, \infty)^{N_k}} \exp(-\frac{1}{2}\gamma_{kj} \mathbf{n}^T \mathbf{R}_k \mathbf{n}) d\mathbf{n} \quad (4.3)$$

is well-defined. For *Tikhonov regularization*, where  $\mathbf{R}_k = \mathbf{I}_{N_k}$  holds, we have a closed form expression for it, namely

$$C_{kj} = \left( \frac{\pi}{2\gamma_{kj}} \right)^{\frac{N_k}{2}}.$$

In *minimal first differences regularization* with zero boundary conditions the regularization matrix is given by

$$\mathbf{R}_k = \mathbf{H}_k^T \mathbf{H}_k \quad \text{with} \quad \mathbf{H}_k = \begin{pmatrix} -1 & & & \\ 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \\ & & & 1 \end{pmatrix}.$$

For *Twomey regularization* with eliminated zero boundary conditions we have

$$\mathbf{R}_k = \mathbf{H}_k^T \mathbf{H}_k \quad \text{with} \quad \mathbf{H}_k = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}.$$

Here  $\mathbf{R}_k$  is a positive definite tridiagonal matrix. For the latter two regularization methods  $C_{kj}$  must be computed numerically.

Remember that the container *SolutionSets* stores reconstructions from at most 3 discretization levels. We let the index  $i$  run through all discretization levels in *SolutionSets* and the index  $j$  through all residual parameters captured in the  $i$ -th level. Then with Bayes' rule the posterior model probabilities are

$$p(N_k, \gamma_{kt} | \mathbf{e}) = \frac{p(\mathbf{e} | N_k, \gamma_{kt}) p(N_k, \gamma_{kt})}{\sum_i \sum_j p(\mathbf{e} | N_i, \gamma_{ij}) p(N_i, \gamma_{ij})}$$

where with (4.1)-(4.3) we have

$$\begin{aligned} & p(\mathbf{e} | N_i, \gamma_{ij}) \\ &= \int_{\mathbb{R}^{N_i}} p(\mathbf{e}, \mathbf{n} | N_i, \gamma_{ij}) d\mathbf{n} \\ &= \int_{\mathbb{R}^{N_i}} p(\mathbf{e} | \mathbf{n}, N_i) p(\mathbf{n} | N_i, \gamma_{ij}) d\mathbf{n} \\ &= \int_{[0, \infty)^{N_i}} B^{-1} C_{ij}^{-1} \exp(-\frac{1}{2} \|\Sigma_{\sigma}^{-\frac{1}{2}} (\mathbf{K}_i \mathbf{n} - \mathbf{e})\|_2^2 - \frac{1}{2} \gamma_{ij} \mathbf{n}^T \mathbf{R}_i \mathbf{n}) d\mathbf{n}, \end{aligned} \tag{4.4}$$

where

$$B = (2\pi)^{\frac{N_i}{2}} |\det(\Sigma_{\sigma})|^{\frac{1}{2}}$$

$$\text{and } C_{ij} = \int_{[0, \infty)^{N_i}} \exp(-\frac{1}{2} \gamma_{ij} \mathbf{n}^T \mathbf{R}_i \mathbf{n}) d\mathbf{n}.$$

We assumed that the model matrix  $\mathbf{K}_i$  and the regularization matrix  $\mathbf{R}_i$  were implicitly given by each discretization level  $N_i$ , i.e. we actually have  $p(\mathbf{e} | N_i, \gamma_{ij}) = p(\mathbf{e} | N_i, \mathbf{K}_i, \mathbf{R}_i, \gamma_{ij})$ . For simplicity of notation these were omitted. Note that the prior model probabilities  $p(N_i, \gamma_{ij})$  are still free.

For the computation of the above integrals of multivariate Gaussian densities over the first quadrant of each model space  $\mathbb{R}^{N_i}$  we applied an effective pseudo-random integration method described in [11] which implements the routines presented in [12] and [13].

## 4.2 Model Selection for Nonnegativity Constraints

We now turn to the prior model probabilities  $p(N_i, \gamma_{ij})$ . As mentioned in the beginning of Section 3 we assume that a  $\gamma_{min} > 0$  and a  $\gamma_{max} < \infty$  exist which give a lower and an upper bound for the regularization parameters  $\gamma$  in order to exclude improper or point-mass priors for the cases  $\gamma = 0$  or  $\gamma = \infty$ . This assumption is independent of the discretization level. We further assume the discretization level to be independent and uniformly distributed. Thus we are taking a *noninformative* prior, and so the prior model probabilities cancel out and do not affect the posterior probabilities.

Now everything is prepared to perform the model selection. To compute integrals of the form

$$\int_{[0, \infty)^N} \exp(-\frac{1}{2} (\mathbf{n}^T \mathbf{H} \mathbf{n} - 2\mathbf{n}^T \mathbf{v} + q)) d\mathbf{n},$$

where  $N$  is the dimension of the square matrix  $\mathbf{H}$ , we apply the method from [11]. It actually can only evaluate integrals of the form

$$\frac{1}{\sqrt{\det(\mathbf{W})(2\pi)^N}} \int_{a_1}^{b_1} \dots \int_{a_N}^{b_N} \exp(-\frac{1}{2} \mathbf{n}^T \mathbf{W}^{-1} \mathbf{n}) d\mathbf{n},$$

where the cases  $a_i = -\infty$  and  $b_i = \infty$  are allowed. So we have to perform a simple affine transformation

using the Cholesky factorization  $\mathbf{H} = \mathbf{U}^T \mathbf{U}$ :

$$\begin{aligned} & \int_{[0,\infty)^N} \exp(-\frac{1}{2}(\mathbf{n}^T \mathbf{H} \mathbf{n} - 2\mathbf{n}^T \mathbf{v} + q)) d\mathbf{n} \\ &= \left( \exp(-\frac{1}{2}(q - \mathbf{v}^T \mathbf{H}^{-1} \mathbf{v})) \sqrt{\det(\mathbf{H}^{-1})(2\pi)^N} \right) \\ & \cdot \frac{1}{\sqrt{\det(\mathbf{H}^{-1})(2\pi)^N}} \int_{\{\mathbf{z} \in \mathbb{R}^N \mid \mathbf{z} \geq -\mathbf{H}^{-1} \mathbf{v}\}} \exp(-\frac{1}{2} \mathbf{z}^T \mathbf{H} \mathbf{z}) d\mathbf{z}. \end{aligned}$$

The model selection algorithm is as follows.

---

**Algorithm 2** Model Selection

---

```

1: get  $S_1, \dots, S_{MaxDisc}$  from SolutionSets
2: get  $A_1, \dots, A_{MaxDisc}$  from ApproxSets
3: get  $P_1, \dots, P_{MaxDisc}$  from PriorSets
4: get  $T_1, \dots, T_{MaxDisc}$  from TauSets
5:  $m_1 = |S_1|, \dots, m_{MaxDisc} = |S_{MaxDisc}|$ 
6:  $m_{total} = \sum_{k=1}^{MaxDisc} m_k$ 
7:  $B = (2\pi)^{\frac{N_l}{2}} |\det(\Sigma_{\sigma})|^{\frac{1}{2}}$ 
8:  $P_{post} = \{\}$ 
9: for  $i = 1$  to  $MaxDisc$  do
10:   for  $j = 1$  to  $m_i$  do
11:      $\mathbf{K}_{ij} = A_i(j)$ 
12:      $\mathbf{R}_{ij} = \frac{1}{\delta^2} P_i(j)$ 
13:      $C_{ij} = \int_{[0,\infty)^{N_i}} \exp(-\frac{1}{2} \mathbf{n}^T \mathbf{R}_{ij} \mathbf{n}) d\mathbf{n}$ 
14:      $M_{ij} = \int_{[0,\infty)^{N_i}} \exp(-\frac{1}{2} \|\Sigma_{\sigma}^{-\frac{1}{2}} (\mathbf{K}_{ij} \mathbf{n} - \mathbf{e}_{real})\|_2^2 - \frac{1}{2} \mathbf{n}^T \mathbf{R}_{ij} \mathbf{n}) d\mathbf{n}$ 
15:      $P_{post} = P_{post} \cup \{M_{ij}/(B \cdot C_{ij})\}$ 
16:   end for
17: end for
18:  $SumP_{post} = \sum_{k=1}^{m_{total}} P_{post}(k)$ 
19: for  $i = 1$  to  $MaxDisc$  do
20:    $P_{post}(i) = P_{post}(i)/SumP_{post}$ 
21: end for
22:  $S_{total} = S_1 \cup \dots \cup S_{MaxDisc}$ 
23: sort  $S_{total}(1), \dots, S_{total}(m_{total})$  according to  $P_{post}(1), \dots, P_{post}(m_{total})$ 

```

---

In the first lines of the model-selection algorithm the containers for computed reconstructions, operator approximation matrices, prior matrices and residual parameters are loaded for each examined discretization level. They store the results of the model-generation algorithm from Section 3.5. In the case of too noisy or improper data it might happen that in the model generation step none of the models can fit the data. Then all containers are empty and the model selection algorithm has to be aborted. For simplicity we assume that the model-generation step was successful.

The double loop in lines 9-17 performs the multidimensional integrations needed in (4.3) and (4.4). In line 15 these integrals are used for the unnormalized posterior probabilities  $p(N_i, \gamma_{ij} | \mathbf{e})$  from (4.4). Note that the prior model probabilities  $p(N_i, \gamma_{ij})$  do not appear in the algorithm, since they are selected to be uniform and thus cancel out in the normalizing step performed in lines 18 - 21. At last all reconstructions are sorted according to their posterior probabilities.

We have to be careful not to forget to normalize the regularization matrices  $P_i(j)$  with the estimated noise level  $\delta^2$  as in line 12 because all statistical computations have to be carried out using the

unnormalized covariance matrix  $\Sigma_{\sigma}$ . For very small noise levels we recommend to skip the model selection step completely due to instabilities in the statistical computations mentioned above. It is sufficient to use only the coarsest model generated with the commonly used value  $\tau = 1.1$  in this case.

## 5 Numerical Results

### 5.1 Simulation of Aerosol Spectroscopy Measurements

We applied our algorithm to a simplified version of problem (1.2), where we assumed that we know the minimal and maximal particle radii  $r_{min}$  and  $r_{max}$ . This led to the integral equation

$$\int_{r_{min}}^{r_{max}} k(r, l) n(r) dr = e(l). \quad (5.1)$$

For the kernel function  $k(r, l)$  from Mie theory we selected  $\text{H}_2\text{O}$  as the material for the scattering particles and air for the medium.

In our simulations we assumed  $r_{min} = 0.01 \mu\text{m}$  and  $r_{max} = 7.0 \mu\text{m}$ . In practice the extinction function can only be measured for a finite number of light wavelengths  $l_1, \dots, l_{N_l}$ . In our simulations we used the grid of 48 wavelengths composed of 8 linearly spaced wavelengths from  $0.6 - 0.8 \mu\text{m}$ , 8 from  $1.1 - 1.3 \mu\text{m}$ , 8 from  $1.6 - 1.8 \mu\text{m}$ , 16 from  $2.1 - 2.5 \mu\text{m}$  and 8 from  $3.1 - 3.3 \mu\text{m}$ . These five intervals were chosen to exclude wavelengths where light absorption by ambient water can occur which distorts the measured extinctions  $e(l)$  heavily. That is, the selected wavelengths cover the so-called *optical window* which is free from this unwanted physical effect.

We generated artificial extinction values  $e(l_i)$  for the selected  $l_1, \dots, l_{N_l}$  by solving the forward problem, which means inserting an original ‘true’ size distributions  $n(r)$  into the integral equation (5.1). To avoid the *inverse crime* we used a very fine grid with 10001 points and the composite Simpson rule to compute the resulting integrals.

In each simulation run we generated a set of 300 noisy extinctions from the artificial true extinction values by adding zero-mean Gaussian noise where the standard deviations were taken to be 30% of the true extinction values  $e(l_i)$ . This means that a vector  $\mathbf{e}$  of noisy extinctions for each single measurement was modeled as

$$(\mathbf{e})_i = e(l_i) + \delta_i \quad \text{with } \delta_i \sim \mathcal{N}(0, (0.3 \cdot e(l_i))^2), \quad i = 1, \dots, N_l.$$

We used the sample means and variances of these 300 artificial noisy extinctions to do inferences about the simulated Gaussian noise.

For the discretization of (5.1) we used a Galerkin collocation method with linear basis functions on an integration grid with  $N_r = 300$  equidistant points. We generated our model spaces by selecting collocation grids as near equidistant subgrids of the integration grid where the number of grid points  $N_{col}$  ranged from 3 (coarsest discretization level) to 50 (finest discretization level). For the collocation grids we set up linearly spaced ‘pre-collocation grids’ with  $N_{col}$  points first and then performed a nearest-neighbor-fitting of their points to the integration grid, such that they became subgrids. Since we are considering size distributions which attain small values at the minimal and maximal radii, we assumed zero boundary conditions. This effectively reduced the number of unknowns  $N$  in each model space from  $N = 3, \dots, 50$  to  $N = 1, \dots, 48$  and—more importantly—prevented the reconstructed size distributions from sheering out at the smallest radius value, which would have been a not reasonable behavior, physically speaking. It was important that the dimension  $N$  of each model space never succeeded the number of measurements  $N_l = 48$ , such that the resulting regression problems were fully or overdetermined.

Let  $r_1, \dots, r_{N_r}$  denote the integration grid points. Let  $\{r_1 = c_1 < \dots < c_{N_{col}} = r_{N_r}\} \subset \{r_1, \dots, r_{N_r}\}$  be a collocation grid. The *triangular basis functions*  $b_k(r)$ ,  $k = 1, \dots, N_{col}$  are the piecewise linear functions on the intervals  $[c_1, c_2], \dots, [c_{N_{col}-1}, c_{N_{col}}]$  which fulfill

$$b_k(c_j) = \delta_{kj}, \text{ for } j = 1, \dots, N_{col}.$$

We approximated the sought-after function  $n(r)$  with the linear combination

$$n(r) = \sum_{k=1}^{N_{col}} n_k b_k(r), \quad (5.2)$$

where the weights  $n_2, \dots, n_{N_{col}-1} \in \mathbb{R}$  are free variables and  $n_1 = n_{N_{col}} = 0$  holds because of the zero boundary conditions.

Inserting (5.2) into (5.1) yields the linear system of equations for the unknown weights

$$\sum_{k=1}^{N_{col}} n_k \int_{r_{min}}^{r_{max}} k(r, l_i) b_k(r) dr = e(l_i), \quad i = 1, \dots, N_l. \quad (5.3)$$

We applied the composite trapezoidal rule with the integration grid  $r_1, \dots, r_{N_r}$  on the integrals defining the coefficients in above linear system. The resulting coefficient matrix is the matrix  $\mathbf{K}_N$  from Section 3.5 which approximates the integral operator from the left-hand side of (5.1).

## 5.2 Numerical Study

We performed a numerical study for our reconstruction algorithm with model size distributions characterized by a low number of parameters. We varied the parameters in domains giving physically reasonable size distributions and generated noise in the same order of magnitude as observed in real experimental FASP measurements. Therefore the numerical results should give good estimates of the quality of the reconstructions compared to real size distributions. In the same simulation runs we compared our algorithm with existing reconstruction methods.

### 5.2.1 Applied Methods

For all inversion methods applied in our numerical study we selected for the priors Tikhonov, minimal first differences and Phillips-Twomey regularization from Section 4.

In our inversion method we set the Morozov safety factor grid to

$$\tau_1 = 0.6, \tau_2 = 0.7, \dots, \tau_{12} = 1.7.$$

We refer to this as the *constrained method* in the following.

To see that the constraints in the constrained method are worth the computational effort, we compared it with its counterpart without constraints, which we call the *unconstrained method*. It performs the same model generation step based on the discrepancy principle with the same Morozov safety factors grid, but the constraints in (3.1) were dropped. The computations for the model selection are much easier here, since the integrals of the multivariate Gaussian distributions over the parameter spaces can be evaluated analytically.

By reducing the grid of Morozov safety factors in the constrained method simply to the classical value  $\tau = 1.1$  we obtained another method participating in our numerical study. We call it the *Morozov method*. The comparison with it shows whether the grid of Morozov safety factors is justified or not.

We also implemented a classical model-selection method for the unconstrained problem which is independent of the prior. Here we compared the three coarsest models where the discrepancy principle was applicable with the *Bayesian Information Criterion* (BIC), which was first introduced in [14]. The model with the lowest BIC-value

$$-2 \left( -\frac{1}{2} N_l \log(2\pi) - \frac{1}{2} \log(\det(\mathbf{\Sigma}_{\sigma})) - \frac{1}{2} \|\mathbf{\Sigma}_{\sigma}^{-\frac{1}{2}} (\mathbf{K}_N \mathbf{n}_{ml} - \mathbf{e})\|_2^2 \right) + N \log(N_l),$$

where  $\mathbf{n}_{ml}$  is the unconstrained maximum-likelihood solution, is selected here. We call this method the *BIC method*.

### 5.2.2 Model Size Distributions

We generated the simulated measurement data vectors  $\mathbf{e}_{true}$  by inserting one of the following three model size distributions adopted from [15] into our integral equation (5.1):

1. *log-normal distribution*

$$n(r) = \frac{A}{\sqrt{2\pi}\sigma r} \exp\left(-\frac{1}{2\sigma^2}(\log(r) - \log(\mu))^2\right) \quad (5.4)$$

with amplitude  $A$ , standard deviation  $\sigma$ , and mean  $\mu$ .

2. *Rosin-Rammler-Sperling-Bennet (RRSB) distribution*

$$n(r) = \frac{AN}{\nu} \left(\frac{r}{\nu}\right)^{N-1} \exp\left(-\left(\frac{r}{\nu}\right)^N\right) \quad (5.5)$$

with amplitude  $A$ , exponent  $N$ , and mean  $\nu$ .

3. *Hedrih distribution*

$$n(r) = \frac{128Ar^3}{3\eta^4} \exp\left(-\frac{4r}{\eta}\right) \quad (5.6)$$

with amplitude  $A$  and mean  $\mu$ .

For each simulated size distribution we set the amplitude to  $A = 10^4$ . We choose the remaining parameters so that the relation

$$n(r_{max}) \leq Tol \quad (5.7)$$

with  $r_{max} = 7.0 \mu\text{m}$  and  $Tol = 10$  was satisfied. This is to be consistent with the assumption, that we can neglect the tails of the distributions and truncate them at the maximal radius  $r_{max}$ . Furthermore we assumed the modal value of the log-normal and RRSB distributions to be greater or equal to  $1.0 \mu\text{m}$  in order to exclude too peaked distributions. For each of the above three model size distributions we looped in our simulations through a set of 100 possible parameters satisfying (5.7).

For the log-normal distributions we first selected for the mean  $\sigma$  a linearly spaced grid with ten points ranging from 0.2 to 0.5, i.e.  $\sigma_k = 0.2 + 0.3\frac{k-1}{9}$ ,  $k = 1, \dots, 10$ . Then we saw after a lengthy calculation that (5.7) is equivalent to

$$r_{max} \exp\left(-\left(-2\sigma^2 \log\left(\frac{\sqrt{2\pi}r_{max}\sigma Tol}{A}\right)\right)^{\frac{1}{2}}\right) \geq \mu.$$

The modal value of the log-normal distribution is  $r_{mod} = \exp(\log(\mu) - \sigma^2)$ , so  $r_{mod} \geq 1.0$  is equivalent to  $\mu \geq 1.0 \exp(\sigma^2)$ . Using the last two inequalities we selected

$$\begin{aligned} \mu_{kj} &= 1.0 \exp(\sigma_k^2) + \frac{j-1}{9} (v_k - 1.0 \exp(\sigma_k^2)) \\ \text{with } v_k &= r_{max} \exp\left(-\left(-2\sigma_k^2 \log\left(\frac{\sqrt{2\pi}r_{max}\sigma_k Tol}{A}\right)\right)^{\frac{1}{2}}\right), \\ k &= 1, \dots, 10, j = 1, \dots, 10. \end{aligned}$$

These are the 100 parameters used for the log-normal distributions.

For the RRSB distributions we took for the exponents  $N$  the integer values  $N_k = k+2$ ,  $k = 1, \dots, 10$ . We computed the auxiliary variables  $p_k$  as the real-valued solutions of the equations

$$p_k \exp(-p_k) = \frac{r_{max} Tol}{AN_k}$$

being greater than one. With some algebra one can see that

$$\nu \leq r_{max} \cdot p_k^{-\frac{1}{N_k}}$$



is then equivalent to (5.7) for the RRSB distribution. The modal value of the RRSB distribution is  $r_{mod} = \nu \left( \frac{N-1}{N} \right)^{\frac{1}{N}}$ , therefore  $r_{mod} \geq 1.0$  is equivalent to  $\nu \geq 1.0 \cdot \left( \frac{N-1}{N} \right)^{-\frac{1}{N}}$ . Using the last two inequalities we selected

$$\nu_{kj} = 1.0 \cdot \left( \frac{N_k - 1}{N_k} \right)^{-\frac{1}{N_k}} + \frac{j-1}{9} \left( r_{max} \cdot p_k^{-\frac{1}{N_k}} - 1.0 \cdot \left( \frac{N_k - 1}{N_k} \right)^{-\frac{1}{N_k}} \right),$$

$$k = 1, \dots, 10, j = 1, \dots, 10.$$

Thus we have 100 parameters for the RRSB distributions.

For the Hedrih distribution we found that (5.7) is equivalent to  $\eta \leq \eta_{max}$  with  $\eta_{max} \approx 2.0566$ . Thus we took for  $\eta$  the values

$$\eta_k = 0.8 + \frac{k-1}{99} (\eta_{max} - 0.8), \quad k = 1, \dots, 100.$$

For each of the three size distribution classes we simulated ten artificial noisy measurement-data vectors  $\mathbf{e}$  as described in Section 5.1 for each of the corresponding 100 parameters. This resulted in total in 1000 single simulated FASP experiments for one model size distribution class.

For every inversion we computed the  $L^2$ -error of the obtained reconstruction relative to the original size distribution and measured the total run time needed for the inversion. The computations were performed on a notebook with a 2.27 GHz CPU and 3.87 GB accessible primary memory.

### 5.2.3 Average $L^2$ -Errors

Log-Normal Distribution			
method	average $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	21.3917	21.8413	23.6893
Morozov	25.9880	26.2520	27.5610
unconstrained	31.1588	34.1079	37.7806
BIC	46.4902	48.3897	50.7711

RRSB Distribution			
method	average $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	18.6192	17.8924	17.6709
Morozov	23.8631	23.2843	23.4669
unconstrained	29.1867	33.2711	37.6596
BIC	74.1802	76.7546	80.4552

Hedrih Distribution			
method	average $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	14.3414	13.2150	12.8981
Morozov	25.6963	24.3244	23.5999
unconstrained	36.7919	36.6560	36.7877
BIC	42.7048	41.8407	41.5387

#### 5.2.4 Average Run Times

Log-Normal Distribution			
method	average run times (s)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	1.3525	1.3987	1.4068
Morozov	0.3136	0.3284	0.3325
unconstrained	0.1581	0.1570	0.1570
BIC	0.0324	0.0324	0.0327

RRSB Distribution			
method	average run times (s)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	1.7258	1.8086	1.8048
Morozov	0.3863	0.4108	0.4093
unconstrained	0.1523	0.1516	0.1512
BIC	0.0357	0.0357	0.0360

Hedrih Distribution			
method	average run times (s)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	1.2522	1.2732	1.2872
Morozov	0.2554	0.2640	0.2670
unconstrained	0.1572	0.1565	0.1561
BIC	0.0315	0.0315	0.0314

### 5.3 Average Model Space Dimensions

Log-Normal Distribution			
method	average model space dimensions		
	Tikhonov	min. first fin. diff.	Twomey
constrained	6.7570	7.0890	7.3670
Morozov	8.6010	9.2220	9.4540
unconstrained	5.2670	5.3580	5.4070
BIC	7.2310	7.2310	7.2310

RRSB Distribution			
method	average model space dimensions		
	Tikhonov	min. first fin. diff.	Twomey
constrained	10.4220	10.9830	10.7100
Morozov	11.6910	12.4760	12.1860
unconstrained	8.7370	9.5090	9.2930
BIC	8.6840	8.6840	8.6840

Hedrih Distribution			
method	average model space dimensions		
	Tikhonov	min. first fin. diff.	Twomey
constrained	6.0970	6.3850	6.6390
Morozov	7.5220	7.9560	8.1180
unconstrained	4.7820	4.7980	4.8200
BIC	7.4290	7.4290	7.4290

### 5.4 Extreme Cases

If the relative error of the reconstruction (compared with the original size distribution) is equal or even greater than 100 percent, we regard the inversion as failed. Note that the inversion methods returned  $\mathbf{n} \equiv 0$  by default if none of the kernel matrices in any of the model spaces would yield a reconstruction. Now we list how many times the inversion methods failed in our test runs. To see how trustworthy the results are we present the worst case  $L^2$  errors as well. Finally we display the worst case run times.

#### 5.4.1 Reconstruction Failures

Log-Normal Distribution			
method	number of $L^2$ -errors $\geq 100$ % (out of 1000)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	0	0	0
Morozov	27	27	27
unconstrained	0	0	0
BIC	20	20	20

RRSB Distribution			
method	number of $L^2$ -errors $\geq 100$ % (out of 1000)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	2	1	1
Morozov	60	60	63
unconstrained	0	0	0
BIC	42	47	47

Hedrih Distribution			
method	number of $L^2$ -errors $\geq 100$ % (out of 1000)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	0	0	0
Morozov	33	33	33
unconstrained	0	0	0
BIC	12	12	12

#### 5.4.2 Worst Case Reconstruction Errors

Log-Normal Distribution			
method	worst case $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	58.7991	58.7264	58.7173
Morozov	525.3771	557.9417	579.4758
unconstrained	70.6745	64.0843	67.1404
BIC	$1.0245 \cdot 10^4$	$1.0246 \cdot 10^4$	$1.0248 \cdot 10^4$

RRSB Distribution			
method	worst case $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	118.8957	114.6170	114.6434
Morozov	389.9993	379.7349	399.7682
unconstrained	85.2596	81.1347	82.4548
BIC	$2.2651 \cdot 10^4$	$2.2656 \cdot 10^4$	$2.2663 \cdot 10^4$

Hedrih Distribution			
method	worst case $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	51.1577	38.2242	35.3611
Morozov	200.3302	224.3174	239.6155
unconstrained	56.6067	56.5325	56.4147
BIC	$1.0273 \cdot 10^4$	$1.0273 \cdot 10^4$	$1.0273 \cdot 10^4$

#### 5.4.3 Worst Case Run Times

Log-Normal Distribution			
method	worst case run times (s)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	6.2403	7.5913	7.0653
Morozov	2.9394	3.3588	4.3115
unconstrained	0.7136	0.7075	0.6234
BIC	0.1126	0.1689	0.1645

RRSB Distribution			
method	worst case run times (s)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	9.6938	10.1606	10.0524
Morozov	2.2293	2.6416	2.3914
unconstrained	0.4096	0.4453	0.4661
BIC	0.0925	0.0896	0.1010

Hedrih Distribution			
method	worst case run times (s)		
	Tikhonov	min. first fin. diff.	Twomey
constrained	2.9746	2.7281	2.7897
Morozov	1.7559	1.8946	2.0438
unconstrained	0.3564	0.4351	0.3860
BIC	0.0556	0.0625	0.0836

## 5.5 Conclusion

The constrained method had the smallest average  $L^2$ -errors and close to zero failure rates. Only for the RRSB distributions were two, one, and one failures out of 1000 inversions recorded for the different priors, respectively. The overall worst case reconstruction error of 118.8957% was only moderately above 100%. It needed the longest run times from all methods, but even the overall worst case run time of 10.9830 seconds was clearly below our thirty-second requirement. The difference of the average  $L^2$ -errors depending on the three priors we applied was not very prominent. For RRSB and Hedrih distributions they seem to decrease by ca. 0.5 to 1% from the Tikhonov to the minimal first finite differences to the Twomey priors, whereas for log-normal distributions the opposite behavior is the case. For the other inversion methods the  $L^2$ -errors behave similarly depending on the priors. Therefore we cannot determine a prior out of the three we used which always yields the smallest average  $L^2$ -error.

For the Morozov method the average  $L^2$ -errors were for log-normal and RRSB distributions about 4 to 6% higher compared to those of the constrained method, but for Hedrih distributions the errors were 11% larger. The average run times represented only about one fifth of those of the constrained method. However, the numbers of failures was significantly higher. For log-normal and Hedrih distributions roughly 3% of all inversions failed, but for RRSB distributions this was up to 6%. The overall worst case  $L^2$ -error of 579.4758% was clearly higher than 100%.

The run times of the unconstrained method were one third to one half of the Morozov method run times. The unconstrained method was the only method without any failures. The overall worst case  $L^2$ -error was a relatively moderate 85.2596%, but the average  $L^2$ -errors were 5 to 12% bigger than the Morozov method  $L^2$ -errors and already 1.5 to 3 times as big as the constrained method  $L^2$ -errors.

The BIC method was by far the fastest one with run times of only a few hundredths of a second, but the average  $L^2$  errors ranging from ca. 40 to 80% were rather poor. The overall worst case  $L^2$ -error was even  $2.2663 \cdot 10^4\%$ .

For practical FASP experiments we conclude that the constrained method performed best, because its average  $L^2$ -errors were smallest, had virtually no failures, and clearly satisfied our thirty-seconds run-time limit even in the worst cases.

## 6 Two-Component Aerosols

In the preceding sections it was assumed that the aerosol particles consist of a known material, and therefore the refractive indices  $m_{part}(l)$  needed to compute the extinction efficiency  $Q_{ext}(m_{med}(l), m_{part}(l), r, l)$  were given exactly as well. But this is not generally the case in real experimental measurements where typically both size distributions and optical properties of scattering particles are unknown. In the ideal case we could set up an additional device for measuring the aerosol refractive indices and perform a two-stage measurement process, where the first step is to retrieve the refractive indices as preparation for the second step of reconstructing the size distribution, but this is not practical. Indeed all measurement techniques for optical properties of aerosol particles demand a pretreatment of the aerosol itself such as vaporizing it into its gas phase or transforming it into a monodisperse aerosol. This would make the FASP too inefficient to be of practical use.

In real applications we simply want to examine some aerosol components of particular interest. Thus we assume that the aerosol to be investigated is a mixture of a small number of known materials, such that only the problem remains to retrieve the volume fractions of these materials in the whole composite aerosol. As an initial explorative step into this general problem we further assume that the aerosol is made up of only two materials.

To compute the refractive indices of composite aerosols from those of their pure components so-called *mixing rules* are used. Some of these are compared in [16]. Let  $m_1 = k_1 + in_1$  and  $m_2 = k_2 + in_2$  be the refractive indices of two aerosol components for a wavelength  $l$  of the incident light. We adopt the most commonly used rule, the *Lorentz-Lorenz rule*. Here the total refractive  $m_{tot} = k_{tot} + in_{tot}$  is obtained from the relation

$$\frac{m_{tot}^2 - 1}{m_{tot}^2 + 2} = f_1 \frac{m_1^2 - 1}{m_1^2 + 2} + f_2 \frac{m_2^2 - 1}{m_2^2 + 2}, \quad (6.1)$$

where  $f_1$  and  $f_2$  are the volume fractions of the components.

Now our new problem is to invert the parameter-dependent integral equation

$$\int_{r_{min}}^{r_{max}} k_p(r, l) n(r) dr = e(l), \quad (6.2)$$

where the sought-after parameter  $p \in [0, 1]$  characterizes the unknown volume fractions. Let  $m_p(l)$  denote the solution  $m_{tot}$  of (6.1) with  $f_1 = p$  and  $f_2 = 1 - p$ . Then the  $p$ -dependent kernel function is given by

$$k_p(r, l) = \pi r^2 Q_{ext}(m_{med}(l), m_p(l), r, l).$$

Mathematically this means that in addition to inverting it we have to identify the “right” integral operator  $K_p$  from the set

$$\left\{ (K_p n)(l) := \int_{r_{min}}^{r_{max}} k_p(r, l) n(r) dr \mid p \in [0, 1] \right\}.$$

We can easily check that  $k_p(r, l)$  depends continuously on  $p$  and therefore so do the discrete approximations  $\mathbf{K}_{k,p}$  to  $K_p$  as well. We again make Assumption 3.8. So by setting

$$\mathbf{K}_p := \Sigma^{-\frac{1}{2}} \mathbf{K}_{k,p} \quad \text{and} \quad \mathbf{r} = \Sigma^{-\frac{1}{2}} (\mathbf{e}_{true} + \boldsymbol{\delta})$$

as in Section 3.4 we obtain the  $p$ -parametrized quadratic programming problem

$$\min_{\mathbf{n} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{K}_p \mathbf{n} - \mathbf{r}\|_2^2 + \frac{1}{2} \gamma \|\mathbf{n}\|_2^2 \quad \text{s.t.} \quad \mathbf{C} \mathbf{n} \leq \mathbf{b}. \quad (6.3)$$

as in Section 2 for the computation of the maximum a posteriori solution.

## 6.1 Fraction Retrieval for two Aerosol Components

For the determination of the parameter  $p$  we modify the adaptive model-generation algorithm from Section 3.5. As a preparation we prove a continuity result.

**Proposition 6.1.** *The minimizer  $\mathbf{n}_p$  of (6.3) for  $\gamma = 0$  depends continuously on the kernel matrix  $\mathbf{K}_p$ .*

*Proof.* Let  $p_1, p_2 \in [0, 1]$  be arbitrary. We write

$$\mathbf{K}_{p_1} =: \mathbf{K} \quad \text{and} \quad \mathbf{K}_{p_2} =: \mathbf{K} + \mathbf{S},$$

hence  $\mathbf{S} = \mathbf{K}_{p_2} - \mathbf{K}_{p_1}$ . From the continuous dependence of  $\mathbf{K}_p$  on  $p$  we have

$$\lim_{p_2 \rightarrow p_1} \mathbf{S} = 0. \quad (6.4)$$

The first-order necessary conditions for the minimizers  $\mathbf{n}_{p_1}$  and  $\mathbf{n}_{p_2}$  of (6.3) for  $p = p_1$  and  $p = p_2$  are given by the relations

$$\mathbf{K}^T \mathbf{K} \mathbf{n}_{p_1} - \mathbf{K}^T \mathbf{r} + \mathbf{C}^T \mathbf{q}_{p_1} = 0 \quad (6.5)$$

$$\text{and } (\mathbf{K}^T \mathbf{K} + \mathbf{K}^T \mathbf{S} + \mathbf{S}^T \mathbf{K} + \mathbf{S}^T \mathbf{S}) \mathbf{n}_{p_2} - (\mathbf{K} + \mathbf{S})^T \mathbf{r} + \mathbf{C}^T \mathbf{q}_{p_2} = 0, \quad (6.6)$$

with vectors  $\mathbf{q}_{p_1} \geq 0$ ,  $\mathbf{q}_{p_2} \geq 0$ . Subtracting (6.5) from (6.6) yields

$$\mathbf{K}^T \mathbf{K} (\mathbf{n}_{p_2} - \mathbf{n}_{p_1}) + (\mathbf{K}^T \mathbf{S} + \mathbf{S}^T \mathbf{K} + \mathbf{S}^T \mathbf{S}) \mathbf{n}_{p_2} - \mathbf{S}^T \mathbf{r} + \mathbf{C}^T (\mathbf{q}_{p_2} - \mathbf{q}_{p_1}) = 0.$$

Forming the scalar product with  $\mathbf{n}_{p_2} - \mathbf{n}_{p_1}$  yields

$$\begin{aligned} & \langle \mathbf{n}_{p_2} - \mathbf{n}_{p_1}, \mathbf{K}^T \mathbf{K} (\mathbf{n}_{p_2} - \mathbf{n}_{p_1}) \rangle + \langle \mathbf{n}_{p_2} - \mathbf{n}_{p_1}, (\mathbf{K}^T \mathbf{S} + \mathbf{S}^T \mathbf{K} + \mathbf{S}^T \mathbf{S}) \mathbf{n}_{p_2} \rangle \\ & - \langle \mathbf{n}_{p_2} - \mathbf{n}_{p_1}, \mathbf{S}^T \mathbf{r} \rangle + \langle \mathbf{C} (\mathbf{n}_{p_2} - \mathbf{n}_{p_1}), \mathbf{q}_{p_2} - \mathbf{q}_{p_1} \rangle = 0. \end{aligned}$$

With (6.4) we obtain in the limit  $p_2 \rightarrow p_1$

$$\lim_{p_2 \rightarrow p_1} \left( \langle \mathbf{n}_{p_2} - \mathbf{n}_{p_1}, \mathbf{K}^T \mathbf{K} (\mathbf{n}_{p_2} - \mathbf{n}_{p_1}) \rangle + \langle \mathbf{C} (\mathbf{n}_{p_2} - \mathbf{n}_{p_1}), \mathbf{q}_{p_2} - \mathbf{q}_{p_1} \rangle \right) = 0.$$

A calculation as in the proof of Lemma 3.1 shows

$$\langle \mathbf{C} (\mathbf{n}_{p_2} - \mathbf{n}_{p_1}), \mathbf{q}_{p_2} - \mathbf{q}_{p_1} \rangle \geq 0,$$

which finally implies

$$\lim_{p_2 \rightarrow p_1} \mathbf{n}_{p_2} = \mathbf{n}_{p_1}.$$

□

From the last proposition we directly obtain an existence result for an optimal  $p$ .

**Corollary 6.2.** *For  $\gamma = 0$  the residual  $\|\mathbf{K}_p \mathbf{n}_p - \mathbf{r}\|_2$  of the minimizer of (6.3) depends continuously on  $p$ , so there exists a  $p \in [0, 1]$  for which it attains its minimal value.* □

Our next step is to find a condition for uniqueness of this minimizer for  $\gamma = 0$ .

**Proposition 6.3.** *Let  $\gamma = 0$  and  $p \in [0, 1]$  be such that  $\|\mathbf{K}_p \mathbf{n}_p - \mathbf{r}\|_2^2$  minimizes all Tikhonov functionals in (6.3) over the parameter range  $[0, 1]$ . Let  $s \in [0, 1]$ ,  $s \neq p$ , be arbitrary. Then if*

$$\langle \mathbf{K}_p \mathbf{n}_p - \mathbf{K}_s \mathbf{n}_s, \mathbf{r} \rangle \neq 0 \quad (6.7)$$

*holds, the minimizing parameter  $p$  is unique.*

*Proof.* The necessary conditions for  $\mathbf{n}_p$  and  $\mathbf{n}_s$  to be a minimizer of (6.3) are given by

$$\mathbf{K}_p^T \mathbf{K}_p \mathbf{n}_p - \mathbf{K}_p^T \mathbf{r} + \mathbf{C}^T \mathbf{q}_p = 0 \quad (6.8)$$

$$\text{and } \mathbf{K}_s^T \mathbf{K}_s \mathbf{n}_s - \mathbf{K}_s^T \mathbf{r} + \mathbf{C}^T \mathbf{q}_s = 0 \quad (6.9)$$

with vectors  $\mathbf{q}_p \geq 0$ ,  $\mathbf{q}_s \geq 0$ . Assume

$$\|\mathbf{K}_p \mathbf{n}_p - \mathbf{r}\|_2^2 = \|\mathbf{K}_s \mathbf{n}_s - \mathbf{r}\|_2^2,$$

which is equivalent to

$$\langle \mathbf{n}_p, \mathbf{K}_p^T \mathbf{K}_p \mathbf{n}_p \rangle - \langle \mathbf{n}_s, \mathbf{K}_s^T \mathbf{K}_s \mathbf{n}_s \rangle = 2 \langle \mathbf{K}_p \mathbf{n}_p - \mathbf{K}_s \mathbf{n}_s, \mathbf{r} \rangle. \quad (6.10)$$

We form the scalar products of (6.8) with  $\mathbf{n}_p$  and of (6.9) with  $\mathbf{n}_s$ . Then forming the difference of the resulting equations gives

$$\langle \mathbf{n}_p, \mathbf{K}_p^T \mathbf{K}_p \mathbf{n}_p \rangle - \langle \mathbf{n}_s, \mathbf{K}_s^T \mathbf{K}_s \mathbf{n}_s \rangle - \langle \mathbf{K}_p \mathbf{n}_p - \mathbf{K}_s \mathbf{n}_s, \mathbf{r} \rangle + \langle \mathbf{C} \mathbf{n}_p, \mathbf{q}_p \rangle - \langle \mathbf{C} \mathbf{n}_s, \mathbf{q}_s \rangle = 0.$$



Inserting (6.10) yields

$$\langle \mathbf{K}_p \mathbf{n}_p - \mathbf{K}_s \mathbf{n}_s, \mathbf{r} \rangle + \langle \mathbf{C} \mathbf{n}_p, \mathbf{q}_p \rangle - \langle \mathbf{C} \mathbf{n}_s, \mathbf{q}_s \rangle = 0.$$

From (3.5) with  $\mathbf{b} = 0$  we conclude  $\langle \mathbf{C} \mathbf{n}_p, \mathbf{q}_p \rangle = 0$  and analogously  $\langle \mathbf{C} \mathbf{n}_s, \mathbf{q}_s \rangle = 0$ . But then

$$\langle \mathbf{K}_p \mathbf{n}_p - \mathbf{K}_s \mathbf{n}_s, \mathbf{r} \rangle = 0,$$

which contradicts (6.7). Thus if (6.7) holds, the minimizing parameter  $p$  must be unique.  $\square$

Condition (6.7) demands that the kernel matrices  $\mathbf{K}_s$  are sufficiently different to  $\mathbf{K}_p$  so that we get distinguishable residuals of the unregularized solutions. If an  $s \in [0, 1]$  happens to exist with  $\mathbf{K}_s = \mathbf{K}_p$ , condition (6.7) cannot be fulfilled. Unfortunately we are not currently able to check this condition a priori.

We conclude this section with investigating how the unregularized residuals behave for moderate noise levels. In the following the superscript  $\delta$  marks solutions of (6.3) for a data vector  $\mathbf{r}$  contaminated with noise.

**Proposition 6.4.** *We assume that condition (6.7) holds for any noise vector satisfying  $0 \leq \|\delta\|_2 \leq \delta$ . Let  $\mathbf{n}_t$  be the minimizer for the true noise-free model, i.e. the parameter  $t \in [0, 1]$  yields the minimal residual  $\|\mathbf{K}_t \mathbf{n}_t - \Sigma^{-\frac{1}{2}} \mathbf{e}_{true}\|_2^2$  over the whole parameter interval  $[0, 1]$ . Let  $p = p(\delta) \in [0, 1]$  be the parameter yielding the minimal unregularized residual  $\|\mathbf{K}_p \mathbf{n}_p^\delta - \Sigma^{-\frac{1}{2}}(\mathbf{e}_{true} + \delta)\|_2^2$  for the noisy data vector  $\Sigma^{-\frac{1}{2}}(\mathbf{e}_{true} + \delta)$ . Then holds  $\lim_{\|\delta\|_2 \rightarrow 0} p(\delta) = t$ .*

*Proof.* To shorten notation we write  $\mathbf{r}_{true} := \Sigma^{-\frac{1}{2}} \mathbf{e}_{true}$  and  $\boldsymbol{\rho} := \Sigma^{-\frac{1}{2}} \delta$ . Let  $\mathbf{n}_t^\delta$  be the minimizer for the parameter  $t$  and the noisy data vector  $\Sigma^{-\frac{1}{2}}(\mathbf{e}_{true} + \delta)$ , i.e.

$$\mathbf{n}_t^\delta = \underset{\mathbf{n} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{K}_t \mathbf{n} - (\mathbf{r}_{true} + \boldsymbol{\rho})\|_2^2 \quad \text{s.t.} \quad \mathbf{C} \mathbf{n} \leq \mathbf{b}.$$

Then we have the first order necessary conditions

$$\mathbf{K}_t^T \mathbf{K}_t \mathbf{n}_t - \mathbf{K}_t^T \mathbf{r}_{true} + \mathbf{C} \mathbf{q}_t = 0 \tag{6.11}$$

$$\mathbf{K}_t^T \mathbf{K}_t \mathbf{n}_t^\delta - \mathbf{K}_t^T (\mathbf{r}_{true} + \boldsymbol{\rho}) + \mathbf{C} \mathbf{q}_t^\delta = 0, \tag{6.12}$$

with vectors  $\mathbf{q}_t^\delta \geq 0$ ,  $\mathbf{q}_t \geq 0$ . Subtracting (6.12) from (6.11) and scalar multiplying the result with  $\mathbf{n}_t - \mathbf{n}_t^\delta$  gives

$$\langle \mathbf{n}_t - \mathbf{n}_t^\delta, \mathbf{K}_t^T \mathbf{K}_t (\mathbf{n}_t - \mathbf{n}_t^\delta) \rangle + \langle \mathbf{n}_t - \mathbf{n}_t^\delta, \mathbf{K}_t^T \boldsymbol{\rho} \rangle + \langle \mathbf{C} (\mathbf{n}_t - \mathbf{n}_t^\delta), \mathbf{q}_t - \mathbf{q}_t^\delta \rangle = 0.$$

As in the proof of Lemma 3.1 this establishes

$$\begin{aligned} \|\mathbf{K}_t (\mathbf{n}_t - \mathbf{n}_t^\delta)\|_2^2 &\leq \langle \mathbf{n}_t^\delta - \mathbf{n}_t, \mathbf{K}_t^T \boldsymbol{\rho} \rangle \\ &\leq \|\mathbf{K}_t (\mathbf{n}_t - \mathbf{n}_t^\delta)\|_2 \|\Sigma^{-\frac{1}{2}} \delta\|_2, \end{aligned}$$

which gives

$$\|\mathbf{K}_t (\mathbf{n}_t - \mathbf{n}_t^\delta)\|_2 = \mathcal{O}(\|\delta\|_2).$$

Now since the parameter  $p$  minimizes the residuals for the noisy vector  $\mathbf{r}_{true} + \boldsymbol{\rho}$  we can estimate

$$\begin{aligned} \|\mathbf{K}_p \mathbf{n}_p^\delta - (\mathbf{r}_{true} + \boldsymbol{\rho})\|_2 &\leq \|\mathbf{K}_t \mathbf{n}_t^\delta - (\mathbf{r}_{true} + \boldsymbol{\rho})\|_2 \\ &\leq \|\mathbf{K}_t \mathbf{n}_t - \mathbf{r}_{true}\|_2 + \|\mathbf{K}_t (\mathbf{n}_t^\delta - \mathbf{n}_t)\|_2 + \|\Sigma^{-\frac{1}{2}} \delta\|_2. \end{aligned}$$

With the previous finding we see that the upper bound in the last inequality converges to the residual  $\|\mathbf{K}_t \mathbf{n}_t - \mathbf{r}_{true}\|_2$  for  $\|\delta\|_2 \rightarrow 0$ . Thus we obtain in the limit

$$\lim_{\|\delta\|_2 \rightarrow 0} \|\mathbf{K}_p \mathbf{n}_p^\delta - (\mathbf{r}_{true} + \boldsymbol{\rho})\|_2 \leq \|\mathbf{K}_t \mathbf{n}_t - \mathbf{r}_{true}\|_2.$$

By definition of  $t$  we have

$$\|\mathbf{K}_t \mathbf{n}_t - \mathbf{r}_{true}\|_2 \leq \lim_{\|\delta\|_2 \rightarrow 0} \|\mathbf{K}_p \mathbf{n}_p^\delta - (\mathbf{r}_{true} + \boldsymbol{\rho})\|_2,$$

so condition (6.7) finally implies  $\lim_{\|\delta\|_2 \rightarrow 0} p(\delta) = t$ .  $\square$

## 6.2 Convergence Analysis

In this section we show that the regularized solutions from the retrieved aerosol fraction converge to the true solution from the true fraction as the noise level approaches zero. This means that we generalize Theorem 3.11 to the case where the underlying true linear operator must be identified from a known set of possible operators.

**Theorem 6.5.** *Under Assumption 3.8, if condition 6.7 is satisfied for all noise vectors  $\delta$  of random variables, then we have for any  $\alpha(\delta)$  with the properties  $\lim_{\delta \rightarrow 0} \alpha(\delta) = 0$  and  $\lim_{\delta \rightarrow 0} \frac{\delta^2}{\alpha(\delta)} = 0$  that  $\lim_{\delta \rightarrow 0} \mathbb{E}(\|\mathbf{n}_p^{\delta, \alpha(\delta)} - \mathbf{n}_t\|_2) = 0$ . Here  $\mathbf{n}_p^{\delta, \alpha(\delta)}$  is the regularized solution for the retrieved fraction parameter  $p = p(\delta)$ .*

*Proof.* We again use the notations  $\mathbf{r}_{true} := \Sigma^{-\frac{1}{2}} \mathbf{e}_{true}$  and  $\boldsymbol{\rho} := \Sigma^{-\frac{1}{2}} \delta$ . Let  $p = p(\delta)$  be the fraction parameter retrieved by minimizing the unregularized residuals. We write

$$\begin{aligned} \mathbf{n}_p^{\delta, \alpha} &:= \operatorname{argmin}_{\mathbf{n} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{K}_p \mathbf{n} - (\mathbf{r}_{true} + \boldsymbol{\rho})\|_2^2 + \frac{1}{2} \alpha \|\mathbf{n}\|_2^2 \quad \text{s.t.} \quad \mathbf{C}\mathbf{n} \leq \mathbf{b}, \\ \mathbf{n}_p^\alpha &:= \operatorname{argmin}_{\mathbf{n} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{K}_p \mathbf{n} - \mathbf{r}_{true}\|_2^2 + \frac{1}{2} \alpha \|\mathbf{n}\|_2^2 \quad \text{s.t.} \quad \mathbf{C}\mathbf{n} \leq \mathbf{b}, \\ \mathbf{n}_p &:= \operatorname{argmin}_{\mathbf{n} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{K}_p \mathbf{n} - \mathbf{r}_{true}\|_2^2 \quad \text{s.t.} \quad \mathbf{C}\mathbf{n} \leq \mathbf{b}, \\ \mathbf{n}_t &:= \operatorname{argmin}_{\mathbf{n} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{K}_t \mathbf{n} - \mathbf{r}_{true}\|_2^2 \quad \text{s.t.} \quad \mathbf{C}\mathbf{n} \leq \mathbf{b}. \end{aligned}$$

Then we have the estimate

$$\mathbb{E}(\|\mathbf{n}_p^{\delta, \alpha} - \mathbf{n}_t\|_2) \leq \mathbb{E}(\|\mathbf{n}_p^{\delta, \alpha} - \mathbf{n}_p^\alpha\|_2) + \mathbb{E}(\|\mathbf{n}_p^\alpha - \mathbf{n}_p\|_2) + \mathbb{E}(\|\mathbf{n}_p - \mathbf{n}_t\|_2).$$

For the first term in the upper bound, the estimate

$$\mathbb{E}(\|\mathbf{n}_p^{\delta, \alpha} - \mathbf{n}_p^\alpha\|_2) = \frac{\mathcal{O}(\delta)}{\alpha^{\frac{1}{2}}}$$

follows from Proposition 3.10. For the second term, Proposition 3.9 gives

$$\lim_{\alpha \rightarrow 0} \mathbb{E}(\|\mathbf{n}_p^\alpha - \mathbf{n}_p\|_2) = 0.$$

Finally, for the third term from Proposition 6.4 follows

$$\lim_{\delta \rightarrow 0} \mathbb{E}(\|\mathbf{n}_p - \mathbf{n}_t\|_2) = 0.$$

This altogether proves our claim. □

## 6.3 Model Generation under Nonnegativity Constraints for Two-Component Aerosols

Proposition 6.3 motivates us to use the unregularized residuals as model generation criterion, which means that we determine those parameters  $s$ , where they are small. In presence of moderate measurement noise in  $\mathbf{e}$  these parameters lie in the vicinity of the unique true parameter  $p$  as was shown in the proof of Proposition 6.4. In the following we discuss the model generation algorithm extended for two-component aerosols. As in Section 3.5 we compute collocation grids with  $N_1 < \dots < N_m$  points and select a grid of Morozov safety factors  $\tau_1 < \dots < \tau_s$ . Furthermore the refractive indices  $k_1(l_1) + in_1(l_1), \dots, k_1(l_{N_i}) + in_1(l_{N_i})$  and  $k_2(l_1) + in_2(l_1), \dots, k_2(l_{N_i}) + in_2(l_{N_i})$  of two pure aerosol components depending on wavelengths  $l_1, \dots, l_{N_i}$  are given.

---

**Algorithm 3** Model Generation for Two-Component Aerosols

---

```
1:  $MaxDisc = 1$ 
2:  $SolutionSets = \{\}$ 
3:  $ApproxSets = \{\}$ 
4:  $PriorSets = \{\}$ 
5:  $MixRatioSets = \{\}$ 
6:  $TauSets = \{\}$ 
7:  $DiscCntr = 0$ 
8: estimate  $\sigma_1^2, \dots, \sigma_{N_l}^2$  from the sample means approximating the standard deviations of  $e_1, \dots, e_{N_l}$ .
9:  $\delta^2 := \max\{\sigma_1^2, \dots, \sigma_{N_l}^2\}$ 
10:  $\Sigma := \delta^{-2} \cdot \text{diag}(\sigma_1^2, \dots, \sigma_{N_l}^2)$ 
11:  $p_i = \frac{i-1}{N_{frac}-1}, i = 1, \dots, N_{frac}$ 
12:  $N_{frac} = 201$ 
13:  $N_{mean} = 5$ 
14:  $I_{min} = \{\}$ 
15: for  $i = 1$  to  $N_{frac}$  do
16:   for  $j = 1$  to  $N_l$  do
17:     compute  $k_{tot}(l_j) + in_{tot}(l_j)$  from  $k_1(l_j) + in_1(l_j)$  and  $k_2(l_j) + in_2(l_j)$  using (6.1) with
        $f_1 = p_i$  and  $f_2 = 1 - p_i$ 
18:   end for
19:   for  $k = 1$  to  $m$  do
20:     compute kernel matrix  $\mathbf{K}_{ik}$  for  $p_i$  and the collocation grid with  $N_k$  points using
        $k_{tot}(l_1) + in_{tot}(l_1), \dots, k_{tot}(l_{N_l}) + in_{tot}(l_{N_l})$ 
21:   end for
22: end for
23: for  $k = 1$  to  $m$  do
24:    $S_k = \{\}$ 
25:    $A_k = \{\}$ 
26:    $P_k = \{\}$ 
27:    $M_k = \{\}$ 
28:    $T_k = \{\}$ 
29:    $R = \{\}$ 
30:    $RM_{min} = \infty$ 
31:   for  $i = 1$  to  $N_{frac}$  do
32:      $\mathbf{n}_{lsqng} = \underset{\mathbf{n} \in \mathbb{R}^{N_k}}{\text{argmin}} \frac{1}{2} \|\Sigma^{-\frac{1}{2}}(\mathbf{K}_{ik}\mathbf{n} - \mathbf{e}_{real})\|_2^2$  s.t.  $\mathbf{n} \geq 0$ 
33:      $R = R \cup \{\|\Sigma^{-\frac{1}{2}}(\mathbf{K}_{ik}\mathbf{n}_{lsqng} - \mathbf{e}_{real})\|_2^2\}$ 
34:   end for
35:   for  $i = 1$  to  $N_{frac} - N_{mean} + 1$  do
36:      $RM = \text{mean}(R(i), R(i+1), \dots, R(i + N_{mean} - 1))$ 
37:     if  $RM < RM_{min}$  then
38:        $RM_{min} = RM$ 
39:        $t_{min} = \{i, i+1, \dots, i + N_{mean} - 1\}$ 
40:     end if
41:   end for
42:    $t_{cur} = \{t_{min}(1), t_{min}(3), t_{min}(5)\}$ 
43:    $N_{cur} = |t_{cur}|$ 
44:   for  $i = 1$  to  $N_{cur}$  do
45:     for  $j = 0$  to  $s$  do
```

---

---

```

46:         if  $R(t_{cur}(i)) < \tau_j N_l \delta^2 \wedge \tau_j N_l \delta^2 < \|\Sigma^{-\frac{1}{2}} e_{real}\|_2^2$  then
47:             compute  $\gamma_{kij}$  such that
48:              $\mathbf{n}_{trial} = \underset{\mathbf{n} \in \mathbb{R}^{N_i}}{\operatorname{argmin}} \frac{1}{2} \|\Sigma^{-\frac{1}{2}} (\mathbf{K}_{t_{cur}(i),k} \mathbf{n} - e_{real})\|_2^2 + \frac{1}{2} \gamma_{kij} \mathbf{n}^T \mathbf{R}_k \mathbf{n}$  s.t.  $\mathbf{n} \geq 0$ 
49:             with  $\|\Sigma^{-\frac{1}{2}} (\mathbf{K}_{t_{cur}(i),k} \mathbf{n}_{trial} - e_{real})\|_2^2 = \tau_j N_l \delta^2$ 
50:         end if
51:         if  $\mathbf{n}_{trial}$  exists then
52:              $S_k = S_k \cup \{\mathbf{n}_{trial}\}$ 
53:              $A_k = A_k \cup \{\mathbf{K}_{t_{cur}(i),k}\}$ 
54:              $P_k = P_k \cup \{\gamma_{kij} \mathbf{R}_k\}$ 
55:              $M_k = M_k \cup \{p_{t_{cur}(i)}\}$ 
56:              $T_k = T_k \cup \{\tau_j\}$ 
57:         end if
58:     end for
59: end for
60: if  $S_k, A_k, P_k, M_k$  and  $T_k$  not empty then
61:      $SolutionSets = SolutionSets \cup \{S_k\}$ 
62:      $ApproxSets = ApproxSets \cup \{A_k\}$ 
63:      $PriorSets = PriorSets \cup \{P_k\}$ 
64:      $MixRatioSets = MixRatioSets \cup \{M_k\}$ 
65:      $TauSets = TauSets \cup \{T_k\}$ 
66:      $DiscCntr = DiscCntr + 1$ 
67: end if
68: if  $DiscCntr == MaxDisc$  then
69:     break
70: end if
71: end for

```

---

In line 11 the aerosol fraction parameter interval  $[0, 1]$  is approximated with a linearly spaced grid. For each discrete aerosol fraction  $p_i$  the approximation  $\mathbf{K}_{ik}$  to the linear operator  $K_{p_i}$  is computed in lines 15 to 22 for all model space orders  $N_k$ .

In line 23 the main loop for the model generation begins. Note that we first run through all model orders from 1 to  $m$  beginning with the coarsest models before we iterate through all aerosol fractions  $p_i$ . This means that we perform the residual-based search strategy motivated in Proposition 6.3 for each model space separately, where we start with the coarsest model and refine it if necessary.

In lines 31-41 the residuals of the unregularized reconstructions are calculated, and a scan to find the minimal mean of  $N_{mean}$  solutions corresponding to successive parameters  $p_i, p_{i+1}, \dots, p_{i+N_{mean}-1}$  is performed. A subset of the indices  $i, i+1, \dots, i+N_{mean}-1$  corresponding to the residuals with minimal mean is selected in line 42. By filtering out some of the models corresponding to the parameters  $p_i, p_{i+1}, \dots, p_{i+N_{mean}-1}$  with small residuals we ensure that the models to be compared are not too similar. The selected indices are used for the actual model generation in lines 44-59. Here we loop through all preselected Morozov safety parameters  $\tau_1, \dots, \tau_s$  and we propose with them the possible residual values  $\tau_j N_l$  for the discrepancy principle. In line 46 it is checked if the discrepancy principle is applicable.

If the model generation step is successful, the obtained reconstructions accompanied by their kernel and regularization matrices and their aerosol fraction and residual parameters are stored in the containers  $S_k, A_k, P_k, M_k$  and  $T_k$  in lines 51-57.

Finally if the model generation is successful for  $MaxDisc$  model spaces, the model generation loop is terminated in line 69.

## 6.4 Model Selection for Two-Component Aerosols under Nonnegativity Constraints

Not only the model generation procedure has to be generalized to the case of a two-component aerosol, but also the model-selection framework presented in Section 4 needs to be generalized as well. Here we are not just comparing models with different model spaces but also with different underlying operators  $K_p$ . Thus prior probabilities are also needed for the parameters  $p$  which determine the linear operators—or more precisely their approximations—to be compared. Let  $k$  label the model dimensions  $N_k$ , Let  $i$  run through the indices for the aerosol-fraction parameters  $p_i$ , where  $i$  depends on  $k$ , and let  $j$  run through all Morozov safety parameters  $\tau_j$  used for the model generation, where  $j$  depends on  $k$  and  $i$ . Then we can compute the model posterior probabilities by

$$p(N_k, \mathbf{K}_{ik}, \gamma_{kij} | e) = \frac{p(e | N_k, \mathbf{K}_{ik}, \gamma_{kij}) p(N_k, \mathbf{K}_{ik}, \gamma_{kij})}{\sum_u \sum_{v(u)} \sum_{w(u,v)} p(e | N_u, \mathbf{K}_{vu}, \gamma_{uvw}) p(N_u, \mathbf{K}_{vu}, \gamma_{uvw})} \quad (6.13)$$

We assume that  $p(\mathbf{K}_{vu})$  and  $p(N_u, \gamma_{uvw})$  are independent and thus

$$p(N_u, \mathbf{K}_{vu}, \gamma_{uvw}) = p(N_u, \gamma_{uvw}) p(\mathbf{K}_{vu}).$$

We select  $p(\mathbf{K}_{vu})$  to be uniform and adopt  $p(N_u, \gamma_{uvw})$  from Section 4.2. This leads to

$$p(N_u, \mathbf{K}_{vu}, \gamma_{uvw}) = \frac{1}{N_{total}}, \quad (6.14)$$

where  $N_{total}$  is the total number of triplets  $(u, v(u), w(u, v))$ .

Then the model-selection algorithm proceeds in the same way as Algorithm 2, so we do not restate here. The differences to Section 4.2 are that we have already set  $MaxDisc = 1$  in the model generation step and that the single container  $A_1$  stores kernel matrices approximating different operators  $K_{p_i}$ . While in principle the algorithm could continue to compare different discretizations, this only lead to worse results in our simulations. Therefore, once the algorithm finds a discretization level for which reconstructions are at all possible for any of the safety factors, we stop the refinement and simply focus on the problem of identifying the volume fraction.

## 7 Numerical Results for Two-Component Aerosols

### 7.1 Numerical Study

We conducted a numerical study of our inversion algorithm with almost the same settings as the last section but extended for the retrieval of volume fractions of a two-component aerosol. We used the same wavelength grid as in Sections 5.1 and simulated the same model size distributions as in Section 5.2.2. We selected air as ambient medium as well. We extended the grid of Morozov safety parameters to

$$\tau_1 = 0.5, \tau_2 = 0.6, \dots, \tau_{16} = 2.0.$$

If when running through all model spaces none of these safety factors yielded a solution, we performed in this extreme case an another run of the model generation step using a second grid of safety factors given by

$$\tau_1 = 2.5, \tau_2 = 3.0, \dots, \tau_6 = 5.0.$$

This time we did not just simulate an original aerosol consisting purely of  $H_2O$  but instead generated with (6.1) refractive indices of  $H_2O$  and CsI mixtures for the scattering particles. Here the volume fractions of  $H_2O$  ranged through a set of preselected percentages, namely

$$0, 11, 22, 33, 44, 56, 67, 78, 89 \text{ and } 100.$$

For each of the 100 parameters for the log-normal, RRSB or Hedrih distributions we also now have the above 10 fractions. This results in a total of 1000 cases to simulate.

As preparation to run Algorithm 3 we computed the kernel matrices depending on the water volume fraction parameter  $p \in [0, 1]$  for  $p_i = \frac{i-1}{100}$ ,  $i = 1, \dots, 101$  and interpolated each kernel matrix

entry with a cubic spline on a linearly spaced grid with 201 points covering  $[0, 1]$  to increase further the resolution in  $p$ . Thus we have  $N_{frac} = 201$  in Algorithm 3.

Another important difference to Section 5.1 is that the noise level was taken to be only 5% of the true measurement values  $e(l_i)$  instead of 30%. Thus the noisy measurement data vector  $\mathbf{e}$  was here modeled with

$$(\mathbf{e})_i = e(l_i) + \delta_i \quad \text{with} \quad \delta_i \sim \mathcal{N}(0, (0.05 \cdot e(l_i))^2), \quad i = 1, \dots, N_l.$$

We had to take this lower value because the problem of retrieving the aerosol fractions additionally to the size distributions is much more ill-posed than simply reconstructing the size distribution when the scattering material is known.

To investigate the quality of the reconstructions we computed their  $L^2$ -errors relative to the original size distribution. We list them separately for each of the ten original water fractions. We proceed this way for all of our simulation results.

Furthermore we determined the deviations of the reconstructed water volume fractions from the original ones, e.g. when the original fraction was 22% and  $p_{recon} \in [0, 1]$  the retrieved fraction parameter, we calculated the deviation by  $|22 - 100 \cdot p_{recon}|$ %. This showed us how well one can investigate the unknown two-component aerosol only from FASP measurements using our extended inversion algorithm.

We also report how often the inversions failed. There were two main reasons for inversion failures: the first when the relative  $L^2$  was greater than or equal to 100%, the second when the fraction deviation was greater than or equal to 50%. In both cases the reconstruction cannot give any reasonable information about the true size distribution and the true scattering material anymore. Note that in our simulations we returned by default  $\mathbf{n} \equiv 0$  and  $p_{recon} = 0.5$  when no reconstruction could be found in any of the model spaces. For brevity we only list those original fractions where inversion failures occurred

Finally we list the average and worst case inversion run times over all 1000 simulations.

### 7.1.1 Average $L^2$ -Errors

Log-Normal Distribution			
original water volume percent	average $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	33.5018	33.3939	33.6631
11 %	30.0691	30.1394	31.1483
22 %	28.8329	28.5679	29.6568
33 %	24.9249	24.9295	25.8386
44 %	23.4809	23.4363	24.6004
56 %	22.6134	22.2024	23.2278
67 %	20.4546	20.2646	20.3727
78 %	18.9590	18.7913	19.3874
89 %	18.5494	18.7126	18.9227
100 %	17.9452	17.7468	18.6161

RRSB Distribution			
original water volume percent	average $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	38.4939	37.8096	36.4143
11 %	31.7809	31.2368	30.5024
22 %	30.4269	29.9565	29.5171
33 %	27.7925	27.4651	26.2830
44 %	24.2137	24.1607	23.3831
56 %	21.5396	21.8053	21.3910
67 %	19.5181	19.8905	19.5506
78 %	16.9927	16.7956	16.7368
89 %	14.4005	14.4332	14.2989
100 %	12.0441	11.5948	11.6963

Hedrih Distribution			
original water volume percent	average $L^2$ -errors (%)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	17.2492	17.0990	16.5398
11 %	17.5805	17.5246	17.1164
22 %	17.3665	17.1217	16.5678
33 %	15.0919	15.0406	14.6545
44 %	16.9073	16.8601	16.8571
56 %	15.5153	15.5509	15.2833
67 %	14.5155	14.4545	14.5169
78 %	16.7670	16.7708	16.6434
89 %	13.3015	13.3173	13.2861
100 %	11.3083	11.2952	11.4170

### 7.1.2 Average Water Fraction Deviation

Log-Normal Distribution			
original water volume percent	average water fraction deviation (%)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	10.7150	10.7750	10.9350
11 %	7.3700	7.4400	7.5000
22 %	6.1750	6.0450	6.1050
33 %	4.3700	4.3400	4.4800
44 %	3.9550	3.9750	4.0350
56 %	3.2000	3.1900	3.3600
67 %	2.6050	2.5650	2.5850
78 %	2.2350	2.2150	2.1950
89 %	2.2000	2.2300	2.2100
100 %	1.4150	1.2950	1.4150

RRSB Distribution			
original water volume percent	average water fraction deviation (%)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	6.5800	6.6800	6.6600
11 %	5.3000	5.2900	5.2400
22 %	4.7100	4.6700	4.6300
33 %	3.7300	3.7800	3.8000
44 %	3.7300	3.6900	3.6500
56 %	3.1650	3.1950	3.1450
67 %	2.3650	2.3850	2.3450
78 %	1.8050	1.8550	1.8950
89 %	1.3750	1.4150	1.3850
100 %	0.4650	0.4050	0.4050



Hedrih Distribution			
original water volume percent	average water fraction deviation (%)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	5.0800	5.0000	5.0600
11 %	6.3050	6.3150	6.1850
22 %	4.2600	4.4500	4.1000
33 %	4.4350	4.4550	4.5350
44 %	3.2100	3.2300	3.0800
56 %	3.3750	3.3950	3.2450
67 %	1.8700	1.9000	2.0000
78 %	2.7750	2.7550	2.7850
89 %	1.9150	1.9150	1.8750
100 %	1.0050	0.9850	1.0250

## 7.2 Average Model Space Dimensions

Log-Normal Distribution		
average model space dimensions		
Tikhonov	min. first fin. diff.	Twomey
7.5550	7.5550	7.5550

RRSB Distribution		
average model space dimensions		
Tikhonov	min. first fin. diff.	Twomey
11.3940	11.3940	11.3940

Hedrih Distribution		
average model space dimensions		
Tikhonov	min. first fin. diff.	Twomey
6.3330	6.3330	6.3330

## 7.3 Extreme Cases

When the deviation of the retrieved aerosol fraction from the true one exceeded 50% or the  $L^2$ -error between reconstruction and true solution was bigger than 100% we had to regard the reconstruction as failed. We now list when these failures occurred. There were no failed simulations with the Hedrih distribution.

### 7.3.1 Reconstruction Failures

### 7.3.2 Water-Fraction Retrieval Failure

Log-Normal Distribution			
original water volume percent	number of $L^2$ -errors $\geq 100$ % (out of 100)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	1	1	1
11 %	1	1	1
44 %	0	0	1

RRSB Distribution			
original water volume percent	number of $L^2$ -errors $\geq 100$ % (out of 100)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	8	8	7
11 %	1	1	1
22 %	1	1	1
33 %	1	1	1
44 %	1	0	0
67 %	1	1	1

Log-Normal Distribution			
original water volume percent	number of deviations $\geq 50$ % (out of 100)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	3	3	3
11 %	3	3	3
22 %	1	1	1

RRSB Distribution			
original water volume percent	number of deviations $\geq 50$ % (out of 100)		
	Tikhonov	min. first fin. diff.	Twomey
0 %	1	1	1

### 7.3.3 Average and Worst-Case Run Times

Log-Normal Distribution			
run times (s)			
	Tikhonov	min. first fin. diff.	Twomey
average	1.9700	2.0189	2.0453
worst case	8.5281	7.7737	7.8988

## 7.4 Conclusion

A common trend in the results is that the average  $L^2$ -error decreases with increasing original water volume fraction. For log-normal distributions it averages around 31% for water fractions ranging from 0 to 22% and for RRSB distributions it decreases from around 37 to 30% for the same fractions. This poor behavior can also be seen in the numbers of reconstruction failures, which only occurred for

RRSB Distribution			
run times (s)			
	Tikhonov	min. first fin. diff.	Twomey
average	2.4997	2.5488	2.5536
worst case	16.6366	16.6889	16.5122

Hedrih Distribution			
run times (s)			
	Tikhonov	min. first fin. diff.	Twomey
average	1.4207	1.4429	1.4629
worst case	4.7495	4.9839	5.0766

log-normal and RRSB distributions and mostly for water fractions below or equal to 44%. For higher fractions ranging from 56 to 100% the average  $L^2$  was always below 22% for all three size distribution classes and even improving towards 100%.

The water volume fractions deviations behaved in a similar way. They decreased for increasing original fractions, which means that the quality of the water fraction retrieval was improving towards higher original fractions. Water fraction retrieval failures only happened for log-normal and RRSB distributions, when the original fractions were below or equal to 22%. Again the differences in the deviations depending on the priors were only marginal.

The worst case run times never succeeded our thirty-seconds limit. Even in the extreme cases they stayed below 17 seconds. The average run times ranged from ca. 1.7 to 3.3 seconds.

We can conclude that with the settings made in previous section the analysis of two-component aerosols is possible satisfying our demands on run time and accuracy. The standard deviations of the noise in all single measurements have to be reduced from 30% of the true extinction values to 5% in order to obtain results of comparable quality as in Section 5.2.

## 8 Outlook

The model selection problem, i.e. to select appropriate model spaces  $\mathbb{R}^N$ , can also be treated with *Markov Chain Monte Carlo Methods*. Here the posterior distribution is defined as multidimensional distribution living on all model spaces and sampled by the Monte Carlo method. We plan to compare our methods developed to these.

## 9 Acknowledgement

We thank Prof. Dr. Hans-Josef Allelein and Björn Krupa for providing us with experimental FASP measurement data, which was the guide for our simulated measurement data.

This work is sponsored by the German Federal Ministry of Education and Research (BMBF) under the contract number 02NUK022.  
Responsibility for the content of this report lies with the authors.



## References

- [1] B. A. Krupa, P.-M. Steffen, J. Kobalz, and H.-J. Allelein, "Development and qualification of an aerosol generator for investigations under thermal-hydraulic severe accident boundary conditions," *Proceedings of the 16th International Topical Meeting on Nuclear Reactor Thermal Hydraulics NURETH-16, Chicago, Illinois, USA, 30 Aug 2015 - 4 Sep 2015*, 2015.
- [2] B. A. Krupa, H.-J. Allelein, A. Dreizler, V. Ebert, M. Frank, and D. Steiger, "Development and Qualification of Innovative Measurement Devices for Multi-Component Aerosols and Relative Humidity by Using Extinction Photometry," *Proceedings of Annual Meeting on Nuclear Technology (AMNT14), Frankfurt a. Main Germany, May 6 -8, Vol. 45*, 2014.
- [3] W. Salzmann, "Bedienungsanleitung FASP," *Fraunhofer IPM*, 2011.
- [4] G. Mie, "Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen," *Annalen der Physik*, Vol. 25, 1908.
- [5] Q. Fu and W. Sun, "Mie theory for light scattering by a spherical particle in an absorbing medium," *Applied Optics*, Vol. 40, Issue 9, 2001.
- [6] E. Greenberg and S. Chib, "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, Vol. 12, 1996.
- [7] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Springer Science and Business Media, 1996.
- [8] A. Neubauer, "Tikhonov regularization of ill-posed linear operator equations on closed convex sets," *J. Approx. Theory*, vol. 53, pp. 304–320, 1988.
- [9] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*. Springer, 2005.
- [10] E. I. George, H. Chipman, and R. E. McCulloch, "The Practical Implementation of Bayesian Model Selection," *IMS Lecture Notes*, vol. 38, 2001.
- [11] A. Genz, "Numerical Computation of Multivariate Normal Probabilities," *J. of Computational and Graphical Stat.*, vol. 1, 1992.
- [12] H. Niederreiter, "On a Number-Theoretical Integration Method," *Aequationes Mathematicae*, Vol. 8, 1972.
- [13] R. Cranley and T. Patterson, "Randomization of Number Theoretic Methods for Multiple Integration," *SIAM J Numer Anal*, Vol. 13, 1976.
- [14] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, Vol. 6, Issue 2, 1978.
- [15] N. Riefler and T. Wriedt, "Intercomparison of Inversion Algorithms for Particle-Sizing Using Mie Scattering," *Particle and Particle Systems Characterization*, 2008.
- [16] S. Sharma, P. B. Patel, R. S. Patel, and J. J. Vora, "Density and Comparative Refractive Index Study on Mixing Properties of Binary Liquid Mixtures of Eucalyptol with Hydrocarbons at 303.15, 308.15 and 313.15K," *E-Journal of Chemistry*, Vol. 4, Issue 3, 2006.